SITE-SPECIFIC PROTEIN LABELING USING FARNESYL TRANSFERASE

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mohammad Rashidian

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Professor Mark D. Distefano

September 2013

Acknowledgements

There are several people in my life who have helped shape my character and personality, make the decisions I have made, and have supported me on the different levels of my education. I would like to take a moment to acknowledge those that have been instrumental in my life and career.

Several of my school teachers deserve many thanks for making science fun and interesting, and feeding my curiosity with great explanations and rewarding me with motivating prizes. Those teachers are, Mrs. Mojtahedi, Mrs. Banaee, Mrs. Soofi, Mr. Sodagari and Mr. Mousavi. Additionally, my high school teachers were also responsible for molding my curiosity and encouraging my interest in science, especially Mr. Jahankhah, Mr. Eshaghi and Mr. Khazraee. Many of my undergraduate Professors deserve many thanks as well, Professors Alireza Fattahi, Afshin Shafiee, Hashem Nejat, Mansour Abedini and Ali Seyyedi.

To my advisor, Professor Mark Distefano, whom I owe a great debt. You have always been supportive and encouraging. I'm grateful for your patience, motivation, enthusiasm, and immense knowledge. I'm deeply grateful for the freedom that you gave me to decide my projects and research; for helping me to learn how to think better, and how to write better. I have always appreciated your support for such endeavors and the flexibility to allow me to do them. Moreover, you've taught me about the politics of science, and the realities a scientist faces in every aspect of their daily life. In every group meeting, I learned something from you. You have always provided us with helpful

advice, and, in time, I have found nearly all of them to be worthwhile and helpful. I could not have imagined having a better advisor for my Ph.D. study and I'm honored to have been able to start my career under your guidance.

To Professors Wagner and Taton, who have always been very supportive and encouraging. My special thanks to Professor Wagner for enthusiastically supporting me working on the exciting nanoring project.

To my family: You have done a lot to shape my character. You have always supported me by any means. Without you, I'm sure that I could not accomplish what I have accomplished thus far. I'm very grateful to all of you: Mom, Dad, Masoud, Ali, Sarah, Vahid and my dear wife: Ameneh.

To my friends: you have always helped me to enjoy my life. I'm very grateful to you for being supportive and helpful. I have truly been fortunate throughout my life to be surrounded by incredibly nice, bright, encouraging, and enthusiastic people. All that you have done for me will never be forgotten.

Table of Contents

Chapter 1. Chemoenzymatic Labeling of Proteins: Techniques and Approaches

Mohammad Rashidian, Jonathan K. Dozier, Mark D. Distefano

A critical challenge in modern chemical biology is the site-specific modification of proteins. This is due to the large number of reactive functional groups typically present in the cellular environment. An area of intense research is in developing methods for the site-specific modification of proteins, because it has a wide range of utility in fields such as chemistry, biology and medicine. Site-specific protein modification experiments have been useful for oriented protein immobilization, for studies of naturally-occurring post-translational modifications, for creating antibody-drug conjugates, for the introduction of fluorophores and other small molecules on to proteins, for examining protein structure, folding, dynamics and protein-protein interactions and for the preparation of protein-polymer conjugates. One the most important approaches toward protein labeling is to incorporate bioorthogonal functionalities into proteins at specific sites via enzymatic reactions. The incorporated tags would enable reactions that are chemoselective, whose functional groups are both inert in biological media, and which do not natively occur in proteins or other macromolecules. This review article summarizes the enzymatic strategies, which enables site-specifically functionalizing proteins with a variety of different functional groups. The enzymes covered in this review include formylglycine generating enzyme, sialyl transferase, phosphopantetheinyl transferase, O-GlcNAc post-translational modification, sortagging, transglutaminase, farnesyl transferase, biotin ligase, lipoic acid ligase and N-myristoyl transferase.

Introduction

Site-specific modification of proteins is a challenging problem in modern chemical biology.[1–4] With applications in fields including biology, chemistry, and medicine,[5] developing methods to site-specifically modify proteins is of great importance. Site-specific protein modifications have been used to create antibody-drug conjugates,[6] to study natural post-translational modifications, to introduce fluorophores and other small molecules for biophysical studies, oriented protein immobilization,[7] to prepare protein-polymer conjugates,[8] and to examine the properties of proteins such as folding, dynamics and protein-protein interaction.[9] Protein immobilization is an important first step for many biotechnology applications including the construction of protein microarrays and biosensors, immunoassays and protein conjugates used for medical therapies.[10,11] One of the main advantages of site-specific protein modification is that it can be used to create structurally defined covalent linkages between proteins and surfaces, materials or biomolecules.[12] This ensures homogeneous coverage and accessibility to the active site of the protein. These are important factors in studying protein expression and localization, to improve bioavailability and pharmacokinetics of protein-based drugs, in structure-function studies, and in the development of biosensors.[13]

Early methods of protein functionalization exploited the reactivity of either cysteine or lysine residues by reacting the protein with an excess of thiol- or amine-reactive reagents such as maleimides or *N*-hydroxysuccinimidyl esters. In addition to lysine and cysteine modifications, methods have also been developed to chemically modify tyrosine, the N- or C-terminus, and aspartiate or glutamate residues.[14] Transimination[15–17] and periodate

oxidation[4,18] are also among common chemical methods for protein modifications which introduce an aldehyde or a ketone moiety at the N-terminus. Although all of these methods are currently in use and have widespread utility, a variety of alternative approaches have been developed to achieve greater reaction yield, higher selectivity and site-specificity, compatibility with complex biological systems, and improved reaction rates.[12]

One approach to protein labeling is to use reactions that are chemoselective whose functional groups are both inert in biological media and which do not natively occur in proteins or other biological macromolecules. In many cases, a combined strategy of the described modification methods are used, such as taking advantage of the selectivity of the reactive functional groups on proteins (e.g. lysine of the C-terminus) which is used to introduce a new reactive group that then provides superior reaction kinetics for a second, more complex conjugation reaction. Using this strategy, a large excess of the initial reagent can be used to introduce a reactive handle into a protein, followed by a second reaction using stoichiometric quantities of the biomolecule or probe of interest, which can be more costly to obtain.

The discovery and application of green fluorescent protein (GFP) and its derivatives into chemical biology has had a great impact on exploring protein behavior in living cells and has dramatically increased our knowledge of biological processes.[19–21] However, the use of GFP to tag proteins comes with disadvantages; many proteins do not tolerate the fusion with GFP without compromising function or intracellular distribution. An alternative to using GFP is to chemically tag proteins with a fluorescent or affinity label

because this has the advantage of being relatively simple to perform and usually only uses a small molecule when compared to the protein. One of the major disadvantages is that if done non-site specifically, it lacks the precision of a genetically encoded tag.

Another alternative to create site-specific protein conjugates is the incorporation of non-natural amino acids in proteins. The incorporation of non-proteogenic residues allows functionality typically not found in proteins to be introduced and exploited. Several systems for expressing proteins with non-natural residues have been described and recently reviewed.[22] Auxotrophic bacterial hosts are commonly used to incorporate a non-natural analog of an amino acid, and suppressor tRNA methods have been developed which can enable the incorporation of a variety of reactive amino acid analogs such as *p*-azidophenylalanine, *p*-ethynylphenylalanine or p-acetylphenylalanine.[23–26]

The development of new methods for site-specific modification of proteins that function under mild conditions is an area of intense research.[1,3,4,12] Due to the structural sensitivity of proteins and complex cellular environment in which proteins inhabit, the chemical transformations of proteins need to proceed under mild conditions that are compatible with all functional groups present in the cell, which includes such diverse molecules as proteins, nucleic acids, carbohydrates, and various metabolites. The reaction should occur in water or aqueous solutions at or near physiological pH, have rapid kinetics even with low (sub-millimolar) concentrations of reactants, and occur at ambient temperatures. The search for chemical reactions that satisfy these requirements has revealed the utility of a number of key chemical reactions including the Huisgen [3 + 2] cycloaddition (commonly referred to as the "click" reaction),[27–29] the Staudinger ligation,[30] a

photoinducible reaction of an alkene with tetrazole,[31,32] the Diels-Alder reaction,[33–35] an inverse-electron-demand Diels Alder reaction,[36,37] and oxime or hydrazone ligations.[38–40] To overcome many of the aforementioned challenges, chemoenzymatic methods that allow site-specific incorporation of labels into proteins have been developed and are an increasingly intense area of research.[2,41,42] Such methods capitalize on the exquisite selectivity of enzymes to accomplish highly selective reactions. Here we review the most recent enzymatic methods for site-specific protein modification including formylglycine generating enzyme, sialylation, phosphopantetheinyl transferase, *O*-GlcNAc post-translational modification, sortagging, transglutaminase, farnesyl transferase, biotin ligase, lipoic acid ligase and N-myristoyl transferase

Enzymes used for protein modification

Formylglycine generating enzyme (FGE). Formylglycine generating enzyme (FGE) performs a critical posttranslational modification of type I sulfatases by converting a cysteine within the motif CxPxR to the aldehyde-bearing residue formylglycine (FGly).[43–46] By introducing this motif into proteins and subsequently reacting it with FGE, an aldehyde group can be site-specifically added to the protein, which can then be used for covalent modification using complimentary aminooxy- or hydrazide-functionalized reagents (Figure 1.1). Bertozzi and coworkers have refined this strategy by screening FGE's from *Mycobacterium tuberculosis* and *Streptomyces coelicolor* against synthetic peptide substrate libraries and identified new peptide sequences that diverge from the canonical FGE-recognized motif.[44] Based on their study, *Escherichia coli*'s

FGE activity was found to be similarly promiscuous, which enables the use of novel aldehyde tag sequences for *ex vivo* modification of recombinant proteins.

In another study, they introduced the peptide sequence recognized by the endoplasmic reticulum (ER)-resident FGE, into heterologous proteins expressed in mammalian cells. Applying the FGE technique, they were able to site-specifically modify monoclonal antibodies as well as membrane-associated and cytosolic proteins expressed in mammalian systems.[45]

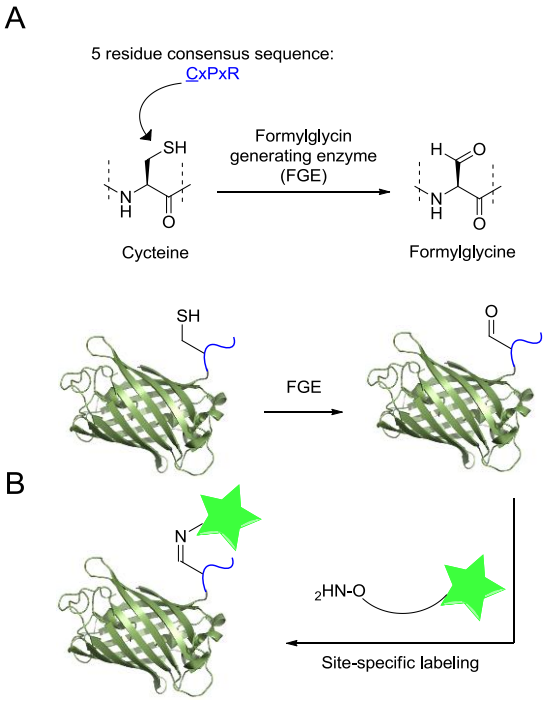


Figure 1.1 Site-specific modification of proteins containing a genetically encoded aldehyde tag incorporated using formylglycine generating enzyme (FGE). (A) FGE oxidizes a specific cysteine to formylglycine within a 5 residue consensus sequence. (B) Site-specific modification of a protein bearing the genetically encoded aldehyde tag followed by chemoselective bioorthogonal oxime ligation reaction to label the protein of interest.

Plasma membrane-associated proteins, similar to secreted proteins, traffic through the secretory pathway and thus can potentially be substrates for the ER-resident FGE. They explored this idea by introducing a 13 residue aldehyde tagging sequence onto the N-terminus of the platelet-derived growth factor receptor (PDGFR) transmembrane (TM) domain. This protein, along with human FGE, were expressed for 48 h in CHO cells. Next, the cells were reacted with biotin-hydrazide followed by Alexafluor488-streptavidin and were analyzed by flow cytometry. Their results showed that CHO cells expressing aldehyde tagged PDGFR-TM had a clear increase in fluorescence compared with cells expressing the unmodified PDGFR-TM protein.[45]

In additional work, a bacterial FGE homolog derived from *Streptomyces coelicolor* was introduced into a mammalian expression vector. A GFP was genetically encoded with a 13 residue aldehyde tag at its N-terminus denoted as Ald_{13}-GFP. Both proteins were expressed in HEK293T cells. After cell lysis and purification, the fusion protein was reacted with biotin-hydrazide and analyzed by non-reducing PAGE and Western blotting. Their results showed that the Ald_{13}-GFP was efficiently labeled with biotin-hydrazide.[45]

Overall, the aldehyde tag offers a versatile method for site-specific modification of membrane associated and cytosolic proteins in mammalian cells. It only requires a peptide sequence 6-13 residues in length that can be varied for the protein modification application. The 6 residue tag is smaller but the 13 residue tag provides higher levels of conversion of Cys to FGly presumably due the greater accessibility to the enzyme active site provided by the longer flanking sequences. Because of its site-selective nature, the

aldehyde tag is promising for the development of new protein-conjugates for research and therapeutic purposes.

Sialylation. The surfaces of many cells, including both prokaryotes and eukaryotes, are decorated with glycan chains. These glycans play important roles in a wide range of biological processes such as cell–cell interaction, protein recognition and small molecule–cell recognition. One important type of glycan modification is the addition of sialic acid to proteins (Figure 1.2). Sialic acid is a monosaccharide with a nine-carbon backbone and is a generic term for the *O*- or *N*-substituted derivatives of neuraminic acid. Sialic acid refers to a wide family of related sugar acids that occur prominently at the terminal position of many eukaryotic surface-exposed glycoconjugates, where they confer important properties upon the resulting cell surface. One of the most abundant and best-studied sialic acid members is *N*-acetylneuraminic acid (Neu5Ac). Besides Neu5Ac, there are numerous naturally occurring variations of sialic acid including substitutions at the C-5 position, or modifications of different hydroxyl groups. Sialic acids are found widely distributed in animal tissues and to a lesser extent in other species, ranging from plants and fungi to yeasts and bacteria, predominantly in glycoproteins and gangliosides.

Sialic acid-containing structures play important roles in cellular recognition and communication. Accordingly, many pathogenic bacteria have evolved the molecular machinery to decorate their cell surfaces with sialic acid to mimic that of their host. This results in the ability to resist the host's innate immune response and also allows them to interact specifically with different host-cell surfaces. Therefore understanding sialylation

will enable a better understanding of how certain pathogens function and could be used to create more effective drugs.

Unfortunately, the study of sialylation is difficult due to the complex structure of sialic acid which presents numerous synthetic challenges. This complexity is due to the hindered tertiary anomeric center and the lack of a neighboring participating group in sialic acids. Using enzymatic methods to create and conjugate sialic acid derivatives is a much more promising method to study sialylation. Thus, a highly active sialyltransferase with broad substrate functionality would be particularly useful in this field.

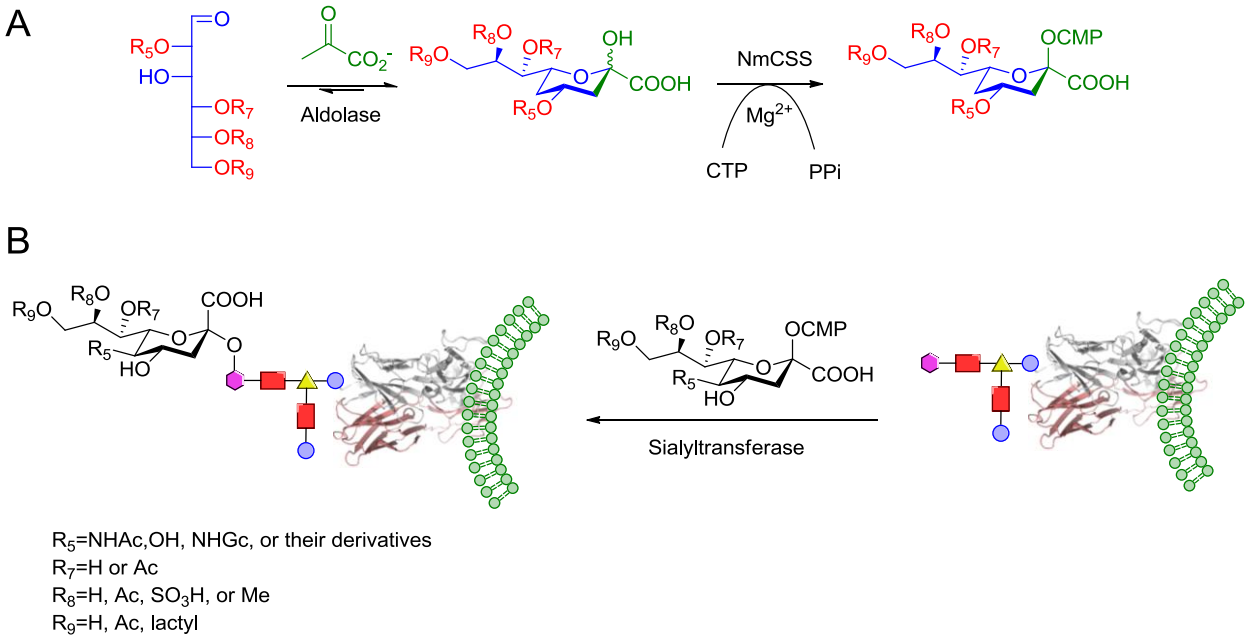


Figure 1.2. (A) Synthesis of sialosides containing sialic acid modifications via a chemoenzymatic approach. (B) Application of sialyltransferase for cell surface labeling using sialic acid derivatives. R_x (X=5-9) represent groups that can contain an azide, an alkyne or other bioorthogonal functional groups.

Chen and coworkers reported the discovery of a water soluble and highly active multifunctional sialyltrasferase from *Pasteurella multocida*.[47] They used a one-pot, three enzyme system for efficient synthesis of a diverse set of sialoside libraries (Figure 1.2).[48] In this system using a recombinant *Escherichia coli* K12 sialic acid aldolase and a recombinant *Neisseria meningitidis* CMP-Sia-synthetase, CMP-Sia derivatives were generated *in situ* from sialic acid precursors. They were able to perform these reactions under mild conditions at 37 ^{o}C, pH of 8.5, over the course of only a few hours; additionally they did not observe significant sialidase or α-2,6-SiaT activity. Overall, the newly discovered SiaT has broad substrate specificity, substantial expression levels, high solubility and is multifunctional. Thus, it should be a powerful tool for synthesizing structurally diverse sialosides to understand their important biological functions.

Phosphopantetheinyl transferase (PPTase). Walsh and coworkers developed a strategy for the site-specific modification of proteins with small molecules using phosphopantetheinyl transferase, Sfp, a posttranslational modification enzyme. Recently, a number of groups have exploited the utility of Sfp to transfer a wide range of molecules, conjugated to the thiol moiety of Coenzyme A (CoA), onto different carrier proteins (Figure 1.3). Sfp and related PPTases have also been used to label proteins, containing the PPTase recognition sequence (peptide carrier protein, PCP, or acyl carrier protein ACP), with biotin[49] or fluorophores, the former of which has been used to noncovalently immobilize arrays of protein onto avidin-coated slides.

In the early work reported by Walsh and coworkers, the target proteins were expressed as fusions to PCP (75-80 amino acids), which were excised from nonribosomal peptide

synthetase.[49] Next Sfp was used to catalyze the covalent attachment of the small molecule phosphopantetheine to a specific serine residue within the PCP. The labeling reaction proceeds with high specificity and efficiency, targeting the PCP-fusion proteins in the cell lysate. The PCP tag was shown to be compatible with various proteins because its fusion onto proteins did not compromise the functional properties of the parent protein. Additionally, Sfp was shown to be promiscuous with respect to the small-molecule probes linked to CoA. This work highlighted the potential of the PCP tag for site-specific protein labeling with small molecules. Moreover, they employed a PCP tag for biotin labeling of proteins and then demonstrated that the resulting biotinylated proteins could be used to create a protein microarray for enzymatic screening. Overall, the relatively small size of the PCP domain (75-80 amino acids), the portability of PCP to various target proteins, the high efficiency of Sfp-catalyzed PCP posttranslational modification and the promiscuity of Sfp with various small-molecule probes makes this approach an attractive alternative for protein labeling.

Having access to promiscuous metabolic pathways can provide an interesting alternative to chemical entry into small molecule engineering *ex vivo*. Burkat and coworkers used the PPTase strategy for protein labeling by harnessing an intrinsic bacterial metabolic pathway.[50,51] They treated *Escherichia coli* cells with a cell-permeable, nonhydrolyzable, fluorescent pantetheine analogue (coumarin-derivative), which can be used by the CoA biosynthetic pathway to synthesize CoA-analogues *ex vivo*. Presumably, *Escherichia coli* internalizes the pantetheine analogue via a sodium-dependent symporter where it undergoes phosphorylation by CoAA. Further processing by CoAD and CoAE results in

stepwise adenylation and further phosphorylation yields the final CoA analogue. This compound is then retained within the cell and is available for downstream transfer into a carrier protein domain. This work showed how cell-penetratable analogues of metabolic precursors could be used *ex vivo* to site-specifically label proteins. The key feature of this approach is the availability of a viable uptake mechanism. Using this strategy, CoA-bound reporter molecules were successfully generated for post-translational protein modification *ex vivo*. This method may be applicable to natural product pathway manipulation as well as applications in conventional molecular and cellular biology. The significant tolerance to structural modification manifested by the enzymes in this pathway should allow delivery of a wide variety of chemical moieties to *ex vivo* processes.

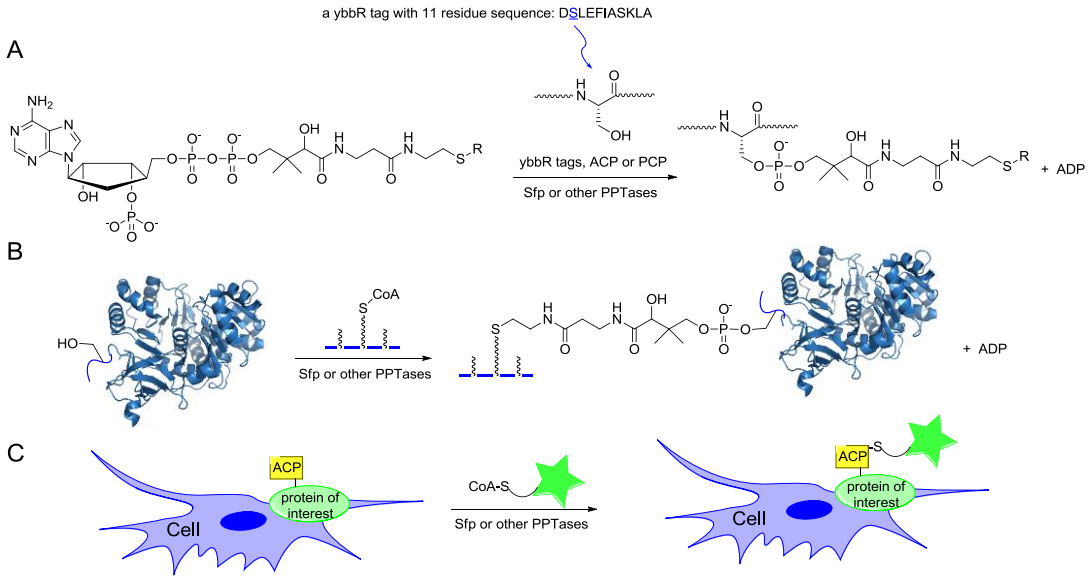


Figure 1.3. (A) Sfp-catalyzed modification of PCP, ACP or ybbR tag at a specific Ser residue by CoA-SR conjugates. (B) Using Sfp or related PPTases for direct site-specific protein immobilization via surface immobilized CoA as a substrate and a protein of interest containing a genetically encoded carrier protein or a ybbR tag. (C) Labeling cell

surface proteins naturally or recombinantly containing ACP, PCP or a ybbR tag using PPTase.

In related work, Johnsson and coworkers used the PPTase strategy to site-specifically label cell surface proteins in eukaryotes and prokaryotes with a fluorescent probe.[52] They expressed a-agglutinin receptor Aga2p as a C-terminal fusion with ACP (Aga2p-ACP) in *Saccharomyces cerevisiae.* Aga2p is known to interact with Aga1p through two disulfide bonds which causes its localization to the cell surface.[53] Incubation of yeast cells that coexpressed Aga1p and Aga2p-ACP with CoA-Cy3 (a fluorescent analog of the natural substrate) and PPTase led to a clear fluorescent signal at the cell membrane. Aga2p-ACP was also modified with a biotinylated CoA analogue, and subsequently incubated with commercially available CdSe quantum dots conjugated to streptavidin. The resulting bright fluorescence of yeast displaying Aga2p-ACP was highly specific. The same results were observed when the experiments were performed on HEK293 cells. In that case, the ACP tag was appended to the N-terminus of a recombinant human G protein-coupled receptor neurokinin-1, one of the most important classes of therapeutic targets.

The size of the PCP or ACP is relatively large (75-80 amino acids), thus Walsh and coworkers tried to identify smaller peptide substrates for Sfp catalyzed reactions. By studying a genomic library of *Bacillus subtilis* by phage display, they identified a series of peptides, the smallest being an 11-residue peptide with the sequence DSLEFIASKLA, as efficient substrates for Sfp phosphopantetheinyl transferase protein labeling by small molecule–CoA conjugates.[54] The peptides were named "ybbR tags" because part of their sequence is derived from the ybbR ORF in the *Bacillus subtilis* genome. Importantly, they also demonstrated that the ybbR tag can be fused to the N- or C-termini of target

proteins or inserted in a flexible loop in the middle of a target protein for site-specific protein labeling by Sfp, highlighting the versatility of this approach.

Overall, enzymatic labeling via Sfp by various small-molecule probes is highly specific and efficient. Sfp has broad substrate specificities regarding the small molecule probes conjugated to CoA, including fluorescent probes such as Alexafluor dyes and fluorescein, affinity probes such as glutathione and biotin, redox probes such as porphyrin and also sugar derivatives of CoA. The PPTase strategy can also be used for direct site-selective covalent protein immobilization on solid supports derivatized with CoA.[55]

O-GlcNAc post-translational modification. The glycosylation of nuclear and cytoplasmic proteins at serine and threonine residues by the addition of a single GlcNAc residue (*O*-GlcNAc) is a widespread posttranslational modification (Figure 1.4). This modification has been found on numerous proteins such as nuclear pore proteins, enzymes and transcription factors. The enzyme catalyzing the *O*-GlcNAc modification of cellular proteins, UDP-Glc-NAc: polypeptidyltransferase (OGTase), is essential for cell viability in mammals and is conserved in a range of organisms from *Arabidopsis thaliana* to humans. Glycosylation levels are typically monitored by immunoblotting with a general *O*-GlcNAc antibody, which detects only a limited number of *O*-GlcNAc-modified proteins and makes it difficult to identify proteins undergoing changes in glycosylation. Bertozzi and coworkers described a chemical strategy directed toward identifying *O*-GlcNAc-modified proteins from living cells or proteins modified *in vitro*.[56] They demonstrated, *in vitro*, that each enzyme in the hexosamine salvage pathway, and the enzymes that affect this dynamic modification (UDP-Glc-NAc:

polypeptidtyltransferase and *O*-GlcNAcase), tolerate structural modifications in which the *N*-acyl side chain of the natural substrate has been functionalized with a bio-orthogonal azide moiety. Cells treated with *N*-azidoacetylglucosamine resulted in the metabolic incorporation of the azidosugar into nuclear and cytoplasmic proteins. These *O*-azidoacetylglucosamine-modified proteins could then be covalently derivatized with various biochemical probes at the site of protein glycosylation using the Staudinger ligation.

The incorporation of GlcNAz onto proteins within cultured human cells was also investigated. Jurkat cells were cultured in the presence of a derivative of GlcNAz for 3 days. Nuclear protein extracts were prepared and reacted with a FLAG-phosphine-based

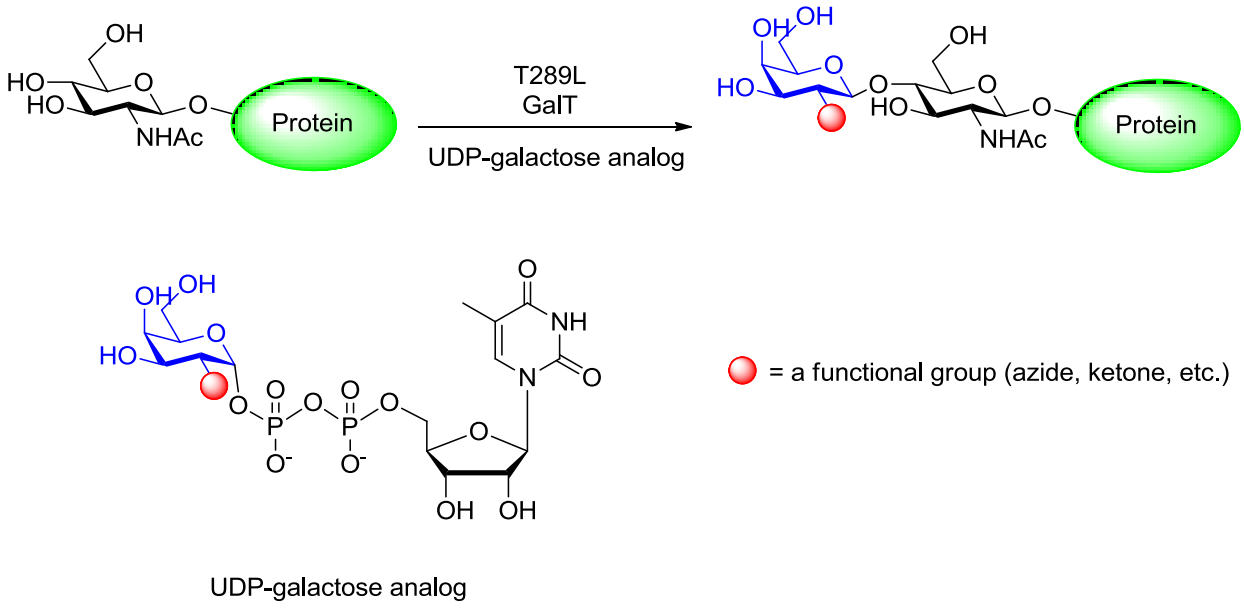


Figure 1.4. Chemoenzymatic labeling of *O*-GlcNAc-containg proteins using a T289L mutant -galactosyl transferase and UDP-galactose analog. The functionalized protein can then be used for chemoselective reactions, such as oxime or hydrazone ligations or azide-alkyne [3+2] cycloaddition.

reagent. Subsequently, FLAG-modified glycoproteins were identified by western blot using a mouse anti-FLAG-mAb conjugated to HRP. As a control, cells that were not treated with the azide modified sugar were analyzed and no signal was detected for this sample. To further validate this approach, they selected a specific protein, the nuclear pore protein p62, which is known to be post-translationally modified with *O*-GlcNAc, and analyzed it from the GlcNAz-treated cells. Using this strategy, protein p62 was successfully labeled with the GlcNAz.

Overall, the exciting result of this approach was that, they could analyze the *O*-GlcNAc-modified proteins in the proteome. This strategy will prove useful for both the identification of *O*-GlcNAc-modified proteins and the elucidation of the specific residues that bear this saccharide. Additionally, this approach could also be used to derivatize *O*-GlcNAc-modified proteins with a variety of other molecules.

Direct enzymatic methods have also been used to detect *O*-GlcNAc posttranslational modifications. Hsieh-Wilson and coworkers reported an enzymatic strategy for a rapid and sensitive detection of *O*-GlcNAc posttranslational modifications.[57–60] Their approach exploits the ability of an engineered mutant of β-1,4-galactosyltransferase to selectively transfer a ketone or azide onto *O*-GlcNAc glycosylated proteins (Figure 1.4) using a suitably functionalized carbohydrate precursor. Next, using a hyrdrazone or oxime ligation or an alkyne-azide cycloaddition reaction employing the transferred ketone or azide, they could site-specifically attach a biotin or a fluorophore to the proteins.

In other work, using this same approach, Hsieh-Wilson and coworkers were able to rapidly visualize proteins that were present at the limits of detection of traditional

methods.[58] As an advantage, this method bypasses the need for radioactive precursors and captures the glycosylated species without perturbing signaling pathways. They examined whether *O*-GlcNAc-modified proteins could be chemoenzymatically tagged and then imaged in cells. They fixed HeLa cells and cultured cortical neurons, permeabilized them, and then labeled them with GlcNAc-azide using the Y289L GalT-engineered enzyme. Next, they reacted the cells with biotin-alkyne or TAMRA-alkyne creating a covalent bond via azide-alkyne cycloaddition. The biotin-treated cells were further incubated with a streptavidin-Alexafluor488 conjugate. Results showed that *O*-GlcNAc-glycosylated proteins were found in both the nucleus and cytoplasm in both type of cells. Moreover, they observed robust staining of proteins along neuronal processes. They labeled nuclear and cytosolic protein fractions from rat forebrain with azide-functionalized *O*-GlcNAc and subsequently modified the material with a TAMRA-alkyne derivative. After enrichment of *O*-GlcNAc proteins by immunoprecipitation using an anti-TAMRA antibody, the precipitates were resolved by 1D or 2D gel electrophoresis, visualized by in-gel fluorescence imaging and then proteolytically digested, and subjected to nano-LC-MS/MS analysis. Those experiments showed minimal nonspecific labeling with the TAMRA-alkyne dye and also resulted in the identification of 213 proteins, among which 67 were previously known and 146 were novel, putative *O*-GlcNAc-modified proteins. Most of the proteins identified participate in neuronal signaling and synaptic function, which suggests important functional roles for *O*-GlcNAc in neuronal communication. Overall, this exciting method enables the direct fluorescence detection of *O*-GlcNAc proteins in gels which in turn facilitates proteomic studies and

extend the reach of existing technologies. This new tool for imaging *O*-GlcNAc-glycosylated proteins also enables the monitoring of expression and dynamics of the modification in cells or tissues.

Sortagging. Recently, the use of the enzyme sortase for labeling proteins has become an area of great interest. Sortase catalyzes a transpeptidase reaction between a specific internal sequence of a protein and an amine group present on the N-terminus of glycine. This method of labeling proteins has been denoted as "Sortagging". One of the most common sortases is Sortase A (SrtA) derived from *Staphylococcus aureus*[61] which recognizes the pentapeptide sequence LPXTG and cleaves the bond between the threonine and glycine, creating a thioester bond between a SrtA-derived cysteine residue and the substrate. This covalent intermediate then undergoes a subsequent aminolysis reaction with the N-terminus of an oligoglycine, thus creating a new covalent bond between the two substrates (Figure 1.5).[62] Additional work has shown that sortagging can be carried out *ex vivo* and cell surface proteins expressing the LPXTG motif can be fluorescently labeled using a small fluorophore. Popp et al. showed that they could take HEK 293T cells expressing CD40L fused with GFP and an LPETG tag, and incubate them with sortase and a TAMRA labeled oligoglycine. They found that such proteins could be labeled *ex vivo*.[63] This type of *ex vivo* labeling can be performed using a wide variety of cell types as well.

Tanaka et al. added a sortase recognition tag, LPETGG to osteoclast differentiation factor (ODF), which is expressed as a cell membrane protein and transfected it into HEK 293T cells. They showed that upon incubating the cells with sortase and biotin connected to a

three-glycine oligomer, they were able to biotinylate on the surface of the cell. They expanded on this initial observation by demonstratng that this approach can work in multiple cell types, including CHO and HeLa cells, and that large molecules, such as eGFP containing a five glycine N-terminus, can be conjugated to the protein on the cells surface.[64] Strijbis et al. showed that labeling can be done in both *Saccharomyces cerievsie* and in the lumen of the ER or cytosol of the HEK293T cells.[65]

Recently, significant progress has been made in creating protein conjugates using SrtA to facilitate covalent attachment. Harrenga et. al. were able to covalently link antibody Fab fragments to a small molecule fluorophore using SrtA at μM concentrations. In this case, they linked the light chain of L19 Fab fragment, which is specific for fibronectinED-B and is a promising target for cancer therapies, to a small peptide containing the fluorophore DY-647. Additionally their work showed that ε-amino groups on lysines can also react causing undesirable protein aggregates. They found that these side reactions were dependent on both the position of the residue within the tertiary structure and on the pH of the reaction.[66] Other examples of creating protein-molecule conjugates include

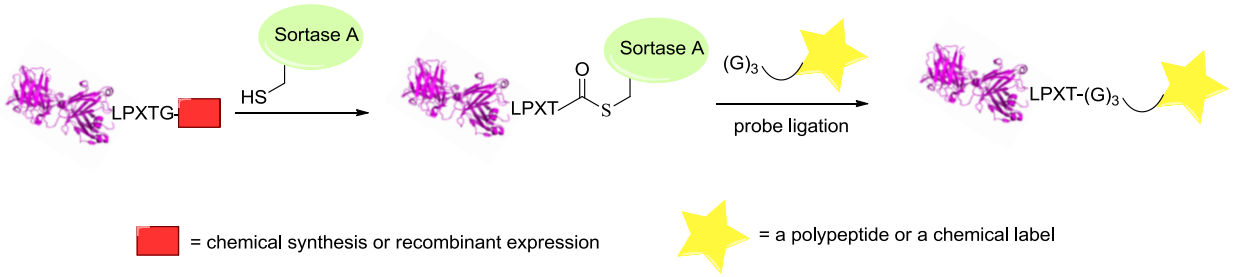


Figure 1.5. Schematic representation of sortase-mediated ligation. First, sortase A recognizes a LPXTG sequence within a polypeptide and forms a covalent acyl-enzyme intermediate. Next, the thioester intermediate is attacked by the N-terminal amino group of a $(G)_3$-containing probe and forms a native peptide bond between the two.

using sortagging to site-specifically PEGylate proteins,[67] to create site-specific protein-lipid conjugates,[68] and to construct peptides and glycosylphosphatidylinositol chimeras.[69] Sortase has also been used in peptide synthesis to cyclize peptides to create macrocyclic peptides and glycopeptides.[70]

In addition to small molecule-protein conjugates, other types of protein conjugates have also been designed. Buti et al. were able to immobilize a library of proteins from an *Influenza* virus onto glass plates using SrtA. In particular, they showed that the LPETG motif could be appended onto the C-terminal ends of proteins which could in turn be site-specifically and covalently attached to the surfaces of glass slides that had been derivatized with a triglycine peptide. Once the proteins were immobilized on the glass plate, they were probed with H1N1 antiserum. Bound H1N1 antibodies were then visualized with a secondary antibody coupled to Alexa-647, whose fluorescence intensity could be quantified. By looking at the difference in fluorescence intensities, they were able to discern which proteins from the *influenza virus* were antigens for the H1N1 antiserum. [71] This not only demonstrates that proteins can be efficiently conjugated to surfaces using sortase but also that these protein arrays can then be used for probing specific biological interactions.

Another application of sortagging is for the creation of protein-protein conjugates. Witte et al. showed that N-to-N and C-to-C protein conjugates could be made using a variation of sortagging. By expressing a protein with triglycine N-terminus and then incubating that protein with SrtA and a small LPETGG peptide incorporating either an azide or alkyne, a bioorthogonal handle was added to the N-terminus of proteins. To create N-to-

N protein-protein conjugates they reacted two complimentary handles to prepare the desired dimer. To create C-to-C protein conjugates, the target protein was expressed to contain a C-terminal LPETGG sequence that was then reacted with an azide- or alkyne-functionalized triglycine peptide. Using this method Witte et al. prepared homodimeric N-to-N fusions of ubiquitin vinylmethylester (UbVME). They were able to show that UbVME was able to still bind its target protein, in this case ubiquitin carboxyl-terminal hydrolase isozyme L3 (UCHL3). Because the dimer UbVME was able to bind two UCHL3 proteins, this confirmed that the UbVME retained its biological activity. Similarly they also constructed C-to-C fusions using a heavy chain antibody, VHH, which binds to GFP. They created this protein conjugate as described above and were able to confirm that it retained biological activity based on the ability of the dimer to bind two GFP proteins. This method demonstrates how sortagging can be used to create protein dimers without comprising the biological activity of the proteins being conjugated.[72]

One of the problems with using sortase as a labeling catalyst is that the reaction is reversible; the glycine residue that is released in the first step can act as nucleophile to reform the original species. This means that an excess of nucleophile is needed to drive the reaction toward product formation. This is especially problematic if the nucleophile is costly or labor intensive to prepare. To address this problem, Williamson et al. developed a depsipeptide, where the amide bond between the threonine and the glycine of the LPETG sequence was replaced with an ester. Their hypothesis was that the peptide would still be a substrate for the enzyme but that the alcohol product could not efficiently

undergo the reverse reaction. They used this peptide first to label a small peptide with the sequence GGSEFG and showed they were able to get almost 100% conversion to product using a 1:1 ratio of the two peptides. They were then able to use this method to fluorescently label both human mannose binding protein and mouse pumilio-2 Puf RNA-binding domain at the N-terminus containing a glycine. Significantly, it was possible to achieve quantitative labeling of the protein using only 1.5 equivalents of the corresponding LPETG-dansylated peptide.[73] This strategy overcomes one of the limitations of using sortagging in that proteins can be labeled at their N-termini without having to use excess amounts of the protein to obtain sufficient levels of labeling.

In summary, sortagging offers a versatile and powerful tool for the site-specific labeling and modification of proteins. Sortagging can be accomplished by directly conjugating proteins to surfaces or other biologically relevant molecules or it can be used to attach bioorthogonal functional groups which can then be used to perform a secondary reaction. As research continues to enhance and refine sortagging, its utility will undoubtedly increase.

Transglutaminase. Transglutaminases are a family of enzymes that catalyze acyl transfer reactions between the carboxamide groups of glutamine residues (present in flexible loops in proteins) to a wide variety of unbranched primary amines, commonly the ε-amino group of lysine. Due to their promiscuity with regard to the primary amine substrate, transglutaminases are attractive catalysts for generating protein conjugates (Figure 1.6).[5]

One of the ways in which transglutaminases can be used to create therapeutic proteins is through the preparation of antibody-small molecule conjugates. Jeger et al. were able to use a bacterial transglutaminase to covalently link antibodies generated against tumor antigens with small molecular chelators that can deliver metals for use as imaging agents. Specifically they conjugated anti-L1-CAM mAb chCE7 antibody and a commercially available anti-CD20 antibody, rituximab. Initial experiments to label the antibodies using transglutaminase failed even though both antibodies had glutamine residues which should have reacted with the enzyme. It was not until the authors employed a mutant form of the antibodies that conjugation was observed. When this was investigated, it was concluded that since the mutation changes the ability of the antibody to be glycosylated and since antibody glycosylation caused the loop containing the targeted glutamine to be in a less accessible conformation, the efficiency of modification is highly dependent on the

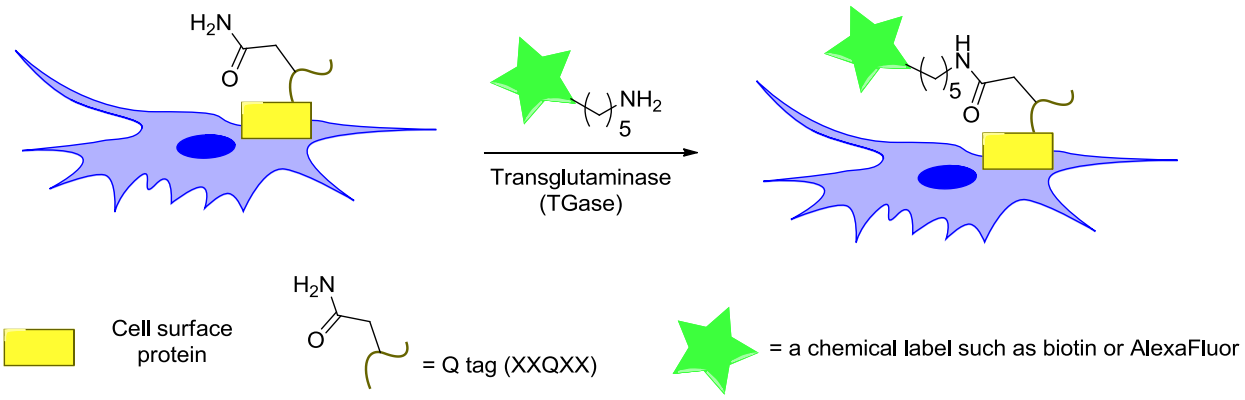


Figure 1.6. Schematic representation of site-specific labeling of proteins using transglutaminase (TGase) catalyzed ligation of cadaverine-functionalized probes (green star) to cell surface proteins recombinantly fused to a Q-tag (yellow rectangle). TGase site-specifically incorporates a probe into the protein by ligating the glutamine of the Q-tag to the amine probe and forms a new amide bond between the two while releasing ammonia into solution.

accessibility of the glutamine which is modified by the enzyme. They were able to follow-up on these findings by treating the antibody with a deglycosylase enzyme, N-glycosidase F, and showed that once the antibodies were deglycosylated, they would then react with transglutaminase. In this way they were able to generate site-specific conjugates of the antibodies.[74]

In addition to creating small molecule-protein conjugates, another area of intense research has been in using transglutaminases to create protein polyethylene glycol (PEG) conjugates. One of the drawbacks to using transglutaminase to PEGylate proteins is that the promiscuity of the enzyme toward glutamine residues makes it difficult to predict the site(s) of modification. This requires using MS method to determine the site of conjugation which can be laborious although progress is being made in this field.[75] Additionally efforts to lipidate proteins using transglutaminase have been reported. Recently Abe et al. reported a new substrate for a *Streptomyces mobaraensis* transglutaminase which can be used to attach lipids to proteins. This substrate consists of an eight amino acid sequence (GGGSLLQG) that is then also attached an alkyl chain. This small peptide-alkyl chain molecule can then act as one of the substrates for the enzyme and be attached to a lysine residue via the enzymatic reaction. In this way a lipid-like modification can be introduced onto various proteins. They used this new substrate to create lipid conjugated eGFPs. The eGFP contained the enzyme recognition sequence MRHKGS, where the lysine acted as the acyl donator, and the lipid substrates ranged between 14 and 18 carbons long. They were then able to show that these protein

conjugates could anchor into the cellular membrane and that the degree on membrane anchoring increased with the length of the alkyl chain.[76] In this way they were able to effectively create protein lipid conjugates using transglutaminase and employ these conjugates to study cellular systems.[76]

Beyond the creation of protein-small molecule conjugates, work has also been performed to create site-specific protein conjugates. Nagamune et al. used a transglutaminase from *Streptomyces mobaraensis* to create eGFP-dihydrofolate reductase(DHFR) protein conjugates. This was accomplished by including a pentaglycine linker at the N-terminus of eGFP and adding a myc tag, which contained the glutaminase recognition sequence, to DHFR. They were then able to show that they could create a covalent linkage between the glycine residue of the N-terminus of eGFP and a glutamine residue near the N-terminus of DHFR. This shows that transglutmaninase can be used to create hetrodimeric protein conjugates. .[77]

Transglutaminase can also be used to covalently attach proteins to a solid support. Tominaga et al. were able to conjugate *Escherichia coli* alkaline phosphatase (AP) to casein-grafted agarose beads. They were able to do this site-specially by adding the TGase recognition tag at the N terminus of AP. Importantly they found that immobilized AP had a very similar specific activity when compared to the soluble enzyme.[78]

Finally, transglutaminases can be used for *ex vivo* labeling as well. Lin et al. used this enzyme to label proteins within living cells. Of particular significance, they used a guinea pig transglutaminase (gpTGase) because it exhibits a high specificity for its glutamine containing protein substrate but also a high promiscuity for the amine substrate. In this

way they could specifically label proteins with what was termed as a Q-tag. Three different amino acid tags were examined including PNPQLPF, PKPQQFM, and GQQQLG, and all were found to be efficiently labeled by gpTGase with a cadaverine-biotin or Alexa568-containing analogue. They showed that this labeling could also be accomplished *ex vivo* by labeling cyan fluorescent protein containing a cell surface targeting sequence using their transglutaminase method. Finally they examined how this method could be used to create an assay for difficult to assay targets. They chose to assay for the presence of the NF-κB homodimer, which can form in the presence of DNA damage but is difficult to easily quanitify. They showed that they could use their transglutaminase method to label NF-κB with a benzophenone spermine analogue, which acts as a photoaffinity reagent, creating a covalent bond between the protein subunits upon photolysis. The presence of the homodimer was easily detected using via SDS-PAGE. This work highlights the utility of transglutaminase for protein labeling since it can be used in a site-specific manner to function in both *in vitro* and *ex vivo* environments.[79] Overall, transglutaminase manifests high potential for use as a site-specific labeling agent. While these enzymes can be somewhat non-selective, ways to overcome this limitation have been developed.

Protein farnesyl transferase (PFTase). In nature, numerous proteins are posttranslationally modified on a sulfur atom present in a cysteine residue at the C-termini of proteins by protein farnesyl transferase (PFTase). PFTase employs farnesyl pyrophosphate (FPP) as its natural substrate to transfer a farnesyl group to the cysteine of a "CaaX-box" tetrapeptide sequence (C is cysteine, a is an aliphatic amino acid, X is one

of a variety of amino acids) (Figure 1.7A&B). The reaction can also be performed *in vitro*. PFTase is promiscuous in the nature of its substrates, in that it can tolerate a variety of CaaX sequences as well as some modifications to the FPP structure. Recently, a number of groups have used this feature to site-specifically modify proteins using analogues of FPP. Various isoprenoid surrogates containing azide,[80] alkyne,[81] aldehyde,[40,82] or 1,3-diamino-phenyl groups have been synthesized and shown to be successfully used by PFTase. The PFTase-catalyzed transfer of synthetic alkyne- or azide-functionalized farnesyl analogues in combination with the Huisgen [3+2] cycloaddition or Staudinger ligation has been used for site-specific fluorophore labeling and oriented immobilization of proteins.[80] Interestingly, CaaX-box sequences such as

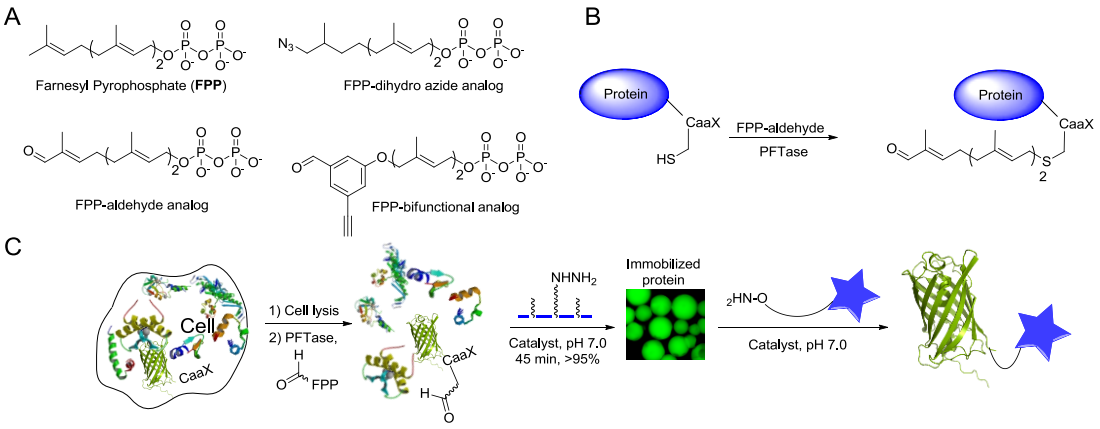


Figure 1.7. (A) Structures of farnesyl diphosphate (FPP), farnesyl dihydro azide diphosphate (FPP-dihydro azide) farnesyl aldehyde diphosphate (FPP-aldehyde) and farnesyl-aldehyde-alkyne diphosphate (FPP-bifunctional) as examples of FPP analogs. (B) Schematic representation of prenylation of a protein containing a CaaX-box positioned at its C-terminus with aldehyde-containing analog (FPP-aldehyde) to yield the prenylated product. (C) Chemoenzymatic site-specific tagging of proteins by FPP-aldehyde using PFTase followed by capture of the aldehyde-functionalized protein in the crude cell lysate via hydrazide functionalized beads. The immobilized protein is then released into solution using aminooxy-containing reagents to yield fluorescently labeled or PEGylated protein.

CVIA, or CVIM can be genetically encoded to the C-termini of many proteins making them efficient substrates for PFTase.

In early work, Poulter and coworkers used this strategy to immobilize protein on solid surfaces.[7] They synthesized alkyne-functionalized FPP and incorporated the alkyne moiety into the proteins (GFP and GST) at their C-termini using PFTase. A glass-slide was functionalized with an azide group and covalent immobilization of the proteins to the slide was subsequently achieved using copper (I)-catalyzed [3+2] cycloaddition. In subsequent work they synthesized a collection of FPP analogues containing either alkyne or azide moieties and analyzed their kinetic parameters (K_M and k_{cat}) as substrates for PFTase.[83] Out of the eleven synthetic analogs, two of them gave steady-state kinetic parameters (K_M and k_{cat}) very similar to those of the natural substrate FPP.

Waldmann and coworkers used the PFTase strategy to modify proteins including H-, N-, and K-Ras GTPases, which are important in cellular signaling and are among the human oncogene products, at their C-termini with a farnesyl group. Subsequently they used thiol functionalized surfaces to react photochemically with the alkenes present in the farnesyl groups introduced into proteins as noted above. In this way, proteins were covalently immobilized in a site-specific manner under mild conditions in about 10 minutes.[84]

In other work, they developed several azide-, biotin- and diene-functionalized analogues of FPP. After kinetic analysis of the enzymatic reactions with PFTase, they concluded that all of these new substrates, except the biotin-containing analogue, can be transferred onto proteins that possess a CaaX box.[85] Subsequently, proteins functionalized with azide

or diene groups were successfully modified by Staudinger ligation or Diels–Alder cycloaddition.

Distefano and coworkers have exploited the PFTase-catalyzed reaction for a number of applications. In early work, they synthesized an analogue of geranyl diphosphate containing an azide group at C-8 and showed that PFTase could transfer that shorter isoprenoid onto a peptide containing a CaaX box.[86] The resulting peptide was further modified by selective Staudinger ligation between the peptidyl azide and a suitable triphenylphosphine derivative; they also found that a longer azide-containing analogue based on FPP could be similarly transferred and reacted.[87] One problem with these azide-containing analogues is that they undergo thermal rearrangements to produce a mixture of isomers due to the allylic azide moiety. Thus, proteins modified with these reagents can potentially produce similar mixtures abrogating one of the advantages of site-selective modification. To avoid this problem, Xu et al. synthesized a dihydroazide analogue based on geraniol that replaces the allylic azide with an aliphatic one and demonstrated that the resulting molecule is an alternative substrate for PFTase.[88] That work was extended by synthesizing a similar analogue based on FPP. In this latter case, the analogue was used to tag GFP appended with a CaaX box with an azide that was subsequently used to immobilize the protein onto alkyne-functionalized agarose beads. Importantly, this two-step enzymatic modification/immobilization sequence was accomplished with both purified protein as well as target present in crude *E. coli* extract at levels as low as 1% of total protein.[80] The versatility of the proteins functionalized with azides incorporated via PFTase was further demonstrated in a subsequent publication where protein-DNA

conjugates were prepared.[89] In that case, a simple GFP-DNA species was first generated by Cu-catalyzed click reaction between GFP-azide and an oligonucleotide equipped with an alkyne. Upon hybridization with a second complementary oligonucleotide containing a Texas Red fluorophore, FRET was observed confirming the production of the desired product. That was extended to the preparation of a more complex protein-DNA tetrahedron composed of four oligonucleotides and four GFP molecules assembled in an analogous manner which was characterized by a combination of electrophoretic and single molecule techniques. The ability to create nanoscale sized protein-based objects with defined architecture highlights the utility of this approach. Recently this approach for preparing protein-DNA assemblies has been extended by using the related strain-promoted azide-alkyne cyclization reaction. By incorporating a dibenzocyclooctyne moiety into the oligonucleotide, a number of proteins were conjugated to DNA including GFP, mCherry, GIP and HIV NC.[90] GFP-mCherry heterodimers were also prepared using this strategy by conjugating complementary alkyne-modified DNA oligonucleotides to the two respective protein-azides followed by hybridization.

In performing the Cu-catalyzed click reaction, it has been noted that lower levels of background reaction are often obtained when the azide component is the one used in excess.[91] Accordingly, Distefano and coworkers prepared an alkyne-containing analogue using a GPP scaffold and showed that it was a good alternative substrate for PFTase.[81] They then showed that more nonspecific immobilization occurred when unfunctionalized GFP was reacted under typical click reaction conditions with alkyne functionalized agarose beads compared to a similar experiment with azide-containing beads. Based on

those results, GFP-alkyne, produced via PFTase-catalyzed incorporation of the alkyne-containing analogue noted above, was labeled with excess Texas Red azide in high (81%) yield. FRET experiments allowed a distance of 37 Å between the GFP and Texas Red fluorophores to be measured which was in good agreement to the distance determined from computer modeling. These experiments highlight the utility of this approach for selective labeling for biophysical experiments. Additional work with a longer alkyne-containing isoprenoid diphosphate based on FPP showed that this new compound was an efficient substrate for both PFTase and the related enzyme protein geranylgeranyltransferase opening up the possibility for transferring larger probes.[92] Conversely, one problem concerning the addition of isoprenoids to proteins is that this modification increases the hydrophobicity of the protein and thereby decreases its solubility. Accordingly, Wollack et al. prepared a more polar PFTase substrate analogue consisting of a single isoprenoid unit linked to an alkyne.[93] While this analogue was incorporated significantly less efficiently (relative to the natural substrate, FPP), it was still possible to achieve essentially quantitative modification of mCherry and Hcp using higher amounts of enzyme. Interestingly, modification of Hcp with the C15-alkyne analogue resulted in a protein that was insoluble; in contrast, the C5-alkyne modified protein was completely soluble highlighting the utility of this less lipophilic modification. In that same report, the authors also demonstrated that the three C-terminal residues following the cysteine residue of the CaaX box could be proteolytically removed by treatment with carboxypeptidase Y after a peptide or protein was modified with the C5 analogue. This means that while PFTase-catalyzed protein modification is not traceless,

the size of the tag remaining in the final protein product can be limited to a single prenylated cysteine residue. This has important implications in the area of protein therapeutics where the antigenicity of any added element may prove deleterious.

Recently, FPP analogues that incorporate aldehyde functionality have also been developed. Rashidian et al. showed that a simple aldehyde-containing analogue of FPP could be used to incorporate an aldehyde into GFP. The resulting aldehyde was used to either immobilize the protein or to label it with a fluorophore via oxime formation.[82] A subsequent study demonstrated that such aldehyde-modified proteins could be prepared and captured from crude extract by incubation with a hydrazide resin. The hydrazone linkage, though stable, is not irreversible and can be converted to a more stable oxime linkage. The authors used this to release, and simultaneously label, the captured protein into the solution by elution with an aryl amine catalyst and an alkoxyamine.[40] Using this strategy, the fluorescently labeled or PEGylated GFP was obtained without prior purification of the starting protein (Figure 1.7C). A PEGylated form of glucose-dependent insulinotropic polypeptide (GIP), a protein with potential as a therapeutic agent for diabetes, was also prepared in an analogous fashion. This approach for protein modification could be particularly useful for large-scale production of protein conjugates for therapeutic or industrial applications.

Biotin ligase. Biotin ligases catalyze the attachment of biotin groups onto proteins and other biomolecules (Figure 1.8). These enzymes possess several useful attributes for protein labeling. They can be used to site-specifically modify proteins by genetic fusion of a ligase recognition domain onto the target protein of interest. Enzymatic labeling of

the protein with biotin allows for the subsequent use of avidin (and related proteins) resulting in the formation of very strong non-covalent conjugates because of the low dissociation constant between biotin and avidin (~10^{-15} M).[94]

A number of different groups have used biotin ligase to site-specifically modify proteins. One common method is to use the *Escherichia coli* biotin ligase, BirA, to attach biotin to proteins. BirA is convenient to use because it attaches biotin onto the lysine of a small, 15-residue, acceptor peptide (AP). This sequence, GLNDIFEAQKIEWHE, was specifically developed as the minimalist structural element that can be efficiently recognized by BirA. The Ting group has adapted this method for preparing a wide range of biotinylated proteins which can be used for a variety of applications. In one study the AP recognition site was attached to a cyan fluorescent protein (CFP) that contains a transmembrane domain (TM) which should traffic the protein to the cell surface of HeLa cells. They found that they could biotinylate the protein using BirA and then conjugate it

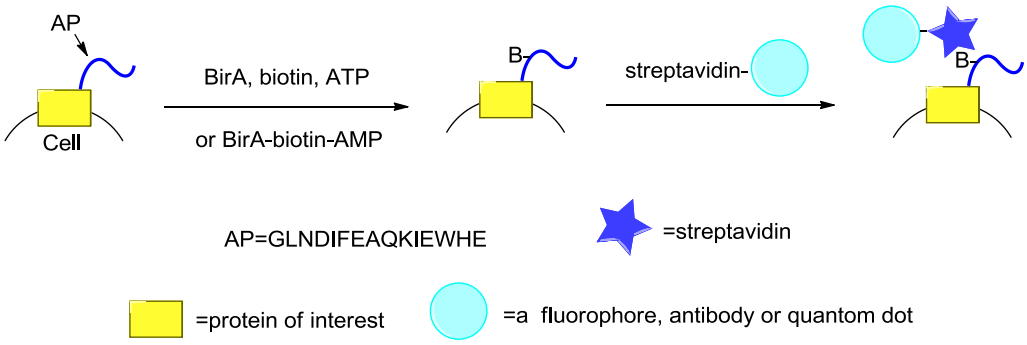


Figure 1.8. Schematic representation of targeting different species such as quantum dots, fluorophors or antibodies to surface proteins in living cells using biotin ligase. The acceptor peptide (AP) tag is genetically encoded either at the C- or N- terminus of a protein. Biotinylation of AP is performed by adding recombinantly expressed biotin ligase (BirA), ATP and biotin (B) to the cell medium. After removing excess biotin by

washing, fluorescently labeled streptavidin is added to visualize the biotinylated surface proteins.
to either a stepavidin-Alexa568 dye or a streptavidin-quantum dot. They determined that the streptavidin-quantum dot gave a better signal then the Alexa568 dye did. They also found that they could label epidermal growth factor receptor using this same method and image it in a similar fashion. They also looked to see if this work would extend to primary cells as well. Hippocampal neurons were transfected an AP tagged GluR2 protein which is a subunit of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptor. They were able to analysis AMPA receptor trafficking by monitoring for fluorescence.[95] More recent work has focused on refining this method so that the quantum dots are smaller and only contain monovalent streptavidin attachments, which can then better label site-specific biotinylated proteins in cells.[96]

Ting and coworkers also explored whether different biotin ligases could catalyze the attachment of biotin analogues that contain alternative functionality suitable for bioorthogonal reactions. A panel of eight different biotin ligases was screened for activity using an array of biotin analogues leading to the discovery that a biotin ligase from *Pyrococcus horikoshii* could act on analogues that contained either an azide or alkyne group. They were then able to label a domain of a biotin acceptor protein with an azide derivative of biotin and conjugate that protein to a FLAG peptide via the Staudinger ligation. In this way they were able to show that biotin ligases could be used to attach bioorthogonal handles to proteins.[97]

Biotin ligase has also been used to prepare modified proteins suitable for studies of protein-protein interactions. In one example, an AP peptide with a high K_M for BirA was

engineered and incorporated into FKBP. The interaction between that protein and a BirA-FRB fusion was then examined. Consistent with the fact that FRB and FKBP interact only in the presence of rapamycin, they observed that labeling of the AP domain by the BirA-FRB fusion occurred only in the presence of both biotin and rapamycin but not in the absence of the rapamycin. These experiments were performed both *in vitro* and *ex vivo* emphasizing the utility of their approach for monitoring protein-protein interaction. To expand the utility of this method they also examined the interaction of Cdc25C with 14-3-3ε phosphoserine/threonine binding protein. Cdc25C was produced as a AP fusion and 14-3-3ε was fused to BirA; both were expressed in HEK cells. Importantly, biotinylation was observed when a wild type-like Cdc25C was used. However, when a non-phosphorylatable Cdc25C mutant was employed, no biotinylation was seen, suggesting that the interaction between Cdc25C and 14-3-3ε is phosphorylation dependent.[98]

Sueda et al. used a biotin ligase from *Sulfolobus tokodaii* to modify proteins for imaging applications.[99] This enzyme biotinylates the lysine residue of a biotin carboxyl carrier protein (BCCP), and interestingly, upon biotinylation, forms a tight non-covalent complex between the ligase and BCCP. Sueda et al. used this unique property to modify proteins by creating fusion proteins between the target of interest and a truncated version of the BCCP protein. They found that the biotin ligase would accept a truncated 69-residue polypeptide derived from BCCP. A fusion was then created between this tag and the bradykinin B2 receptor (B2R), which is a transmembrane protein. They then expressed this protein in HEK293 cells, and incubated these cells in the presence of

Sulfolobus tokodaii biotin ligase that had been labeled previously with four fluorescein molecules. Fluorescent labeling was observed around the cell membrane, confirming that the N-terminus is localized on the outer side of the plasma membrane. In a second experiment, a C-terminal B2R fusion protein with the BCCP tag was expressed in HEK293 cells together with a GFP-biotin ligase fusion protein and were again able to detect fluorescence at the cell membrane. In this case, fluorescent labeling occurred because the C terminus of the B2R protein positioned on the inner side of the membrane was accessible to the GFP-biotin ligase fusion produced intracellularly. As a final remark, it should be noted that while most of the research to date with biotin ligases has focused on *ex vivo* imaging, its site and substrate specificity makes it an attractive tool for creating site-specific protein conjugates for other applications as well. Examples of this will undoubtedly be reported in the future.

Lipoic acid ligase. Lipoic acid ligase catalyzes the acylation of lysine residues of proteins with lipoic acid(Figure 1.9). Significant effort has gone into the development of this enzyme for use in site-specific labeling. Fernández-Suárez et al. designed and synthesized ten different lipoic acid analogues functionalized with either an azide or alkyne and tested them as alternative substrates for *E. coli* lipoic acid ligase (LplA). While all analogues exhibited some degree of activity, they focused on an azide-containing analogue due to its superior activity. Additionally they sought to develop a smaller peptide substrate to minimize the size of the fusion that would have to be used to decrease the possibility of interference with the function of protein under study. By analyzing the structure of E2p, a naturally lipoylated protein, they were able to design a

series of small peptides, which mimicked the β-turn where the lipoylated lysine residue is located on the protein. After further refinement the tag was reduced to just 22 amino acids (LAP). They were then able to show that a CFP-LAP fusion could be labeled with a small fluorophore in human embryonic kidney (HEK) cell lysates. Additionally they added a transmembrane domain to the CFP-LAP protein and were able to show that they could label the protein in living HEK cells and get conjugation using a cyclo-octyne fluorophore, either Cy3 or Alexafluor568. Also a LAP tag was added to the low-density lipoprotein receptor (LDLR) and visualized using this lipoic acid labeling methodology. Finally they tested the cross compatibility of using both lipoic acid labeling and biotin ligase. They transfected one population of cells containing LAP-LDLR and a different population containing an AP fused epidermal growth factor receptor (AP-EGFR). Once the cells were transfected the population were mixed together and subjected to both lipoic acid ligase and biotin ligase conjugation conditions. They saw that only cells containing AP-EGFR showed streptavidin conjugation and cells expressing LAP-LDLR showed only Cy3 labeling. Finally cells were transfected with both types of proteins and it was shown that they could conjugate them with the different bioconjugation techniques.[100]

The use of other lipoic acid analogues for selective protein labeling has also been investigated. Through analysis of the LplA crystal structure, a tryptophan and a glutamate residue in the lipotate binding pocket of the enzyme were identified that when mutated to smaller residues allow the mutant enzymes to accept a wider range of lipoic acid analogues. Using that strategy, it was possible to create a mutant of LplA that accepts lipoate analogues that incorporate fluorescent moieties including Pacific Blue or

7-hydroxycoumarin. Thus, fluorescently labeled proteins could be produced *ex vivo* without having to perform a secondary conjugation reaction.[101,102] These mutant enzymes were also used to incorporate azide analogues containing longer spacers, which were then conjugated *ex vivo* to more red-shifted fluorophores, which have a higher signal to background ratio when compared to the other fluorescent moieties excited at lower

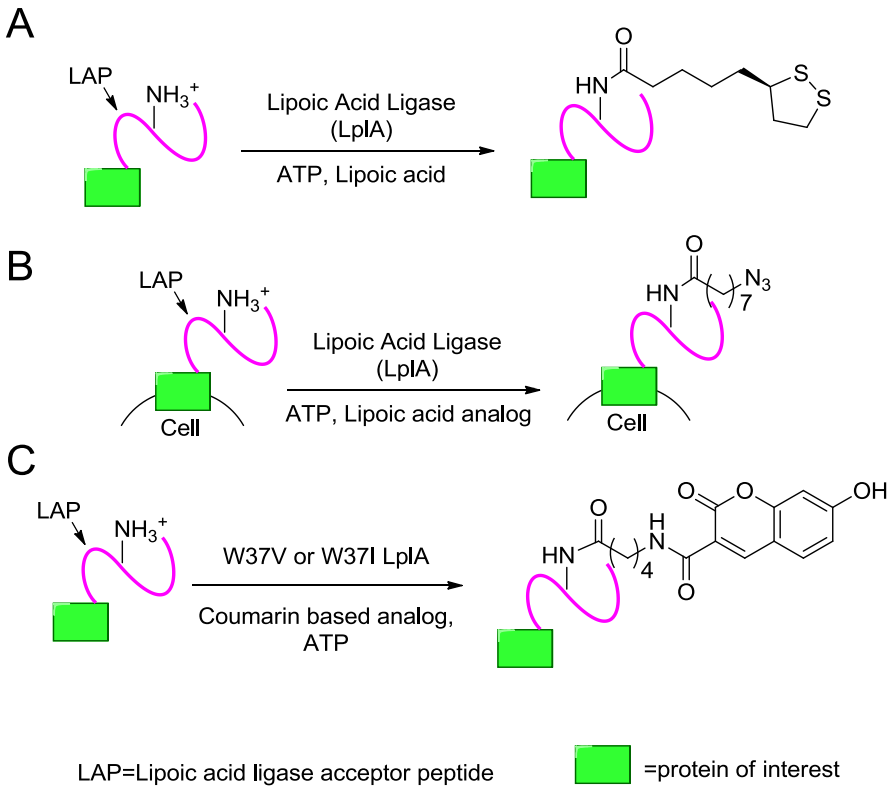


Figure 1.9. Schematic representation of protein labeling using lipoic acid ligase (LplA). (A) Natural ligation reaction catalyzed by LplA. (B) Site-specific labeling of a cell-surface protein performed by fusion of a 22 residue LAP domain on the protein of interest followed by enzymatic transfer of an azide-containing lipoic acid analogue. The azide can be used for further modification via a [3+2] azide-alkyne cycloaddition reaction or Staudinger ligation. (C) Engineered coumarin ligase produced by mutagenesis of LplA can be used to site-specifically modify a protein with a coumarin probe. This can be used to directly incorporate fluorophores to mammalian cells.

wavelengths. Overall, these developments improved the ability to image the locations of proteins within living cells. More recently, Ting and coworkers have applied this lipoic ligase tagging technique to examine protein-protein interactions in experiments similar to those described above with biotin ligase. A lipoic acid peptide tag was created with a high K_M value such that the peptide would only bind when the enzyme was held in close proximity via non-covalent forces. This method is implemented by fusing the lipoic acid ligase enzyme to one target protein and then fusing the peptide tag to a putative binding partner. This technique was employed to study the rapamycin-dependent binding of FKBP and FRB as well as to probe the interaction between the transcription factors Jun and Fos. They also used this technique to study the interaction between neuroligin-1 and PSD-95, which is involved in excitatory synapse formation and maturation. These examples illustrate the versatility of the lipoic acid ligase method for labeling proteins for the analysis of many different types of protein-protein interactions *ex vivo*.[103]

The lipoic acid ligase labeling method has also been adapted to work with lipoic acid tags that contain aryl aldehydes and aryl hydrazines.[104] Ting and coworkers synthesized lipoic acid derivatives containing an aryl aldehyde and an aryl-hydrazine and screened these two analogues against a panel of lipoic acid ligase mutants each containing a mutation at a tryptophan residue at the position 37. They found that a tryptophan to isoleucine mutation significantly increased the extent of conversion when using these alternative substrates compared to that obtained with the wild type enzyme. This mutant enzyme was then used with the aforementioned analogues to label neurexin-1β and then react it with complementary aldehyde- and hydrazine-containing reagents, which contain

either a fluorophore or a chromophore. These reactions were performed *ex vivo* using HEK cells. In a final set of experiments, they used this method to image single low density lipoprotein receptors on the surface of COS7 cells.

Overall lipoic acid ligase is a powerful tool for creating site-specific protein conjugates. As was noted above, in reference to biotin ligase, work using LplA has also centered around the creation of protein-fluorophore conjugates for monitoring cellular processes *ex vivo* and for probing protein-protein interactions. However, this enzyme has significant potential for use in creating other types of protein conjugates including protein-DNA, protein-drug, and protein-protein chimeras. It is likely that such applications will be reported soon.

N-myristoylation. Protein N-myristoylation involves the acyl transfer of myristate from myristoyl-CoA to the amino group of an N-terminal glycine residue of a protein to form an amide bond; The enzyme that performs this modification, N-myristoyl transferase (NMT), recognizes the sequence GXXXS/T where X can be a variety of amino acids and is found in eukaryotes, and very recently it has been observed in prokaryotes as well. Approximately, 0.5% of proteins in eukaryotes are myristoylated. Protein N-myristoylation is thought to occur during protein translation although recent evidence indicates that it can also occur post-translationally. Some important examples of myristoylated proteins include the Src family of tyrosine kinases, HIV-1 matrix protein, HIV-1 Gag and the ADP-ribosylating factors (ARFs). Because of the involvement of these proteins in various diseases, NMT has considerable potential as a therapeutic target.

Tate and coworkers, used the substrate promiscuity of NMT to site-specifically label recombinant proteins with azide and alkyne derivatives of myristate-CoA both *ex vivo* and *in vitro*.[105,106] They used a well-characterized and widely-used NMT cloned from *Candida albicans* to transfer alkyne- and azide-containing myristate analogues that incorporated the bio-orthogonal groups at the distal end of the lipid. First they evaluated the substrate efficiency using a model peptide substrate GLYVSRLFNRLFQKK-NH2 containing a canonical N-terminal myristoylation motif. Incubation of the peptide with either of the two acyl-CoA analogs in the presence of a catalytic quantity of NMT for 2 h resulted in complete transfer of the acyl group to the target peptide.

Next, *Escherichia coli* engineered to co-express NMT and a substrate protein, PfARF1, was used for *ex vivo* analysis. The bacteria cells were transfected with a pair of plasmids encoding the substrate protein and the NMT together with orthogonal antibiotic resistance genes. The cells were incubated with one of the aforementioned analogues and induced by addition of IPTG to express both proteins. Importantly, this strategy

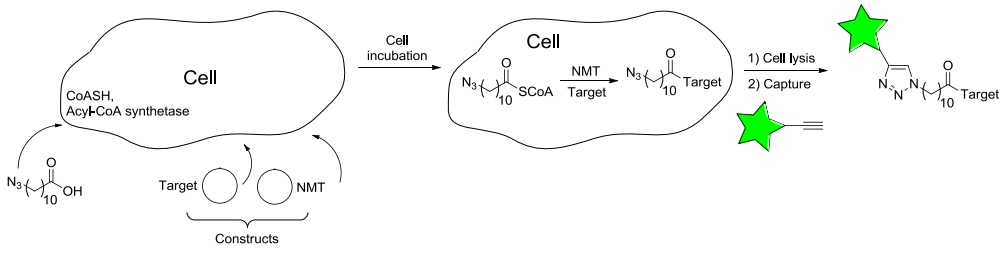


Figure 1.10. Schematic representation of *ex vivo* protein labeling via N-myristoyltransferase (NMT) using an *Escherichia coli* co-expression system. Cells are incubated with a myristic acid analog and plasmids encoding NMT and a recombinant target protein. Within the cells, the analog is converted to its corresponding CoA derivative and incorporated into the desired target protein using NMT. Tagged proteins may then be captured using Staudinger ligation or [3+2] cycloaddition to introduce a secondary label for subsequent applications.

capitalized on endogenous *Escherichia coli* enzymatic activity to convert the myristate analogues to the corresponding -CoA derivatives. After metabolic incorporation, the cells were lysed and reacted with bio-orthogonal biotin reagents to detect the tagged protein including azide-, alkyne- and triarylphosphine-containg compounds. Analysis by SDS-PAGE and subsequent detection with streptavidin-HRP showed that this method could specifically tag PfARF1 when labeling was performed under these conditions within the cell. Overall, this method provides a convenient and potentially general method for N-terminal recombinant protein labeling.

Conclusion.

In summary, a number of enzymes have been studied and employed for site specific labeling of proteins. Each of these enzymes has advantages and disadvantages ranging from substrate specificity to size of the tag added, enzyme kinetics, and site of modification. While all these different factors need to be taken into account, this also gives researchers a number of different options when determining which type of site-specific enzyme to use. Table 1.1 describes a very useful summary of a number of these different parameters for the various enzymes. These methods have been designed so that researchers can now choose between whether they want to label at the N-terminus, C-terminus or somewhere in the middle of the protein. The methods offer a diverse array of labeling approaches that can be performed to modify proteins: everything from modification with small molecules all the way up to modification with a protein. While there has been much progress in way of site specific labeling, there still remains much work to be done so that these processes can be more widely utilized. Many of these

methods have been used primarily in studies of *in vitro* and *ex vivo* processes but have yet to be used in biotechnological processes. The versatility and robustness of many of these reactions make them ideal for creating protein conjugates that can be used in a wide variety of medical and technological applications. Hopefully with these diverse and novel techniques researchers can begin to more fully understand many biological processes and unlock fundamental questions. These new site-specific labeling methods should also help in the development of more powerful and effective drugs against a host of conditions and diseases.

Table 1.1. Kinetic parameters of different enzymes used for protein labeling.

Enzyme	Encoded tag peptide	k_{cat} min^{-1}	K_M µM	k_{cat}/K_M $\mu M^{-1} min^{-1}$	Site of the peptide
Formylglycine generating enzyme	CxPxR or 13-mer LCTPSRGSLFTGR	-	-	-	C or N terminus
Phosphopantetheinyl transferase	ACP, PCP or ybbR tags (11 mer: DSLEFIASKLA; 13 mer: VLDSLEFIASKLA; 17 mer: GSQDVLDSLEFIASKLA)	(for 500 µM ybbR 13 mer and biotin-CoA) 14.7	60.8	0.242	C or N terminus or flexible loops
Sortase	LPXTG	16.2	5,500	0.003	Any section
Transglutaminase	XXQXX	45	7	6	Any section
Farnesyl trasferase	CaaX	(for yPFTase and 2.4 µM CVIA) for FPP: 31.2	1.71	18.2	
		for FPP-aldehyde: 7.8	1.87	4.17	
Biotin ligase	GLNDIFEAQKIEWHE	(for *Escherichia coli* biotin ligase and biotin) 9.6	4.2	2.3	C or N terminus
Lipoic acid ligase	GFEIDKVWYDLDA	2.88	-	-	C or N terminus or flexible loops

Chapter 2. Selective Labeling of Polypeptides Using Protein Farnesyltransferase via Rapid Oxime Ligation

Mohammad Rashidian, Mark D. Distefano

An aldehyde-containing alternative substrate for protein farnesyltransferase was prepared and shown to be enzymatically incorporated into a peptide and a protein. The protein was subsequently immobilized onto aminooxy-functionalized agarose beads or labeled with a fluorophore. This method for protein modification provides an alternative to the commonly employed Cu(I)-catalyzed click reaction.

Introduction. Chemical modification of proteins is important for many applications in biology and biotechnology. Selective functionalization of proteins is challenging because of the large number of reactive functional groups typically present in polypeptides. Our laboratory and others have recently exploited the high specificity of the enzyme, protein farnesyltransferase (PFTase), to site-specifically modify peptides and proteins.[7-80] In nature, PFTase, catalyzes the transfer of a farnesyl isoprenoid group from farnesyl pyrophosphate (FPP, **1**, Figure 2.1) to a sulfur atom present in a cysteine residue. That residue must be located in a tetrapeptide sequence (denoted as a

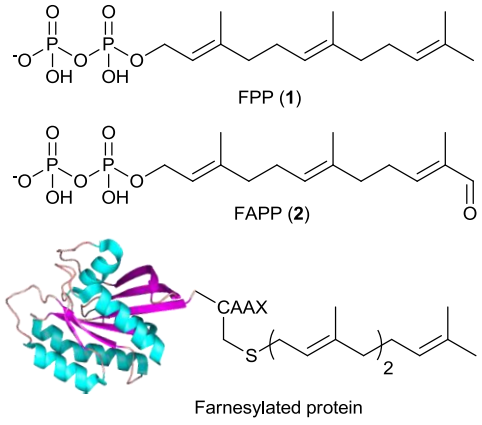


Figure 2.1. Farnesyl pyrophosphate (**1**), farnesyl aldehyde pyrophosphate (FAPP**, 2**) and a farnesylated protein showing a CAAX-box (CVIA) positioned at the C-terminus of a protein.

CAAX-box) positioned at the C-terminus of a protein or peptide to be a PFTase substrate. Interestingly, CAAX-box sequences such as CVIA can be appended to the C-terminus of many proteins rendering them efficient substrates for PFTase. Since PFTase can tolerate many simple modifications to the isoprenoid substrate,[107,108] it can be used to introduce a diverse range of functionality into proteins at their C-termini. Chemoselective reaction with the resulting protein can then be used for a wide range of applications. Although a number of reactions have been developed,[32-11] to date, the Cu(I) catalyzed click reaction has been the most widely used bioorthogonal process.[29] While useful, that reaction employs Cu(I) which is toxic to cells and can erode enzymatic activity.

Ligation reactions between aldehydes or ketones with hydrazine derivatives or alkoxyamines are attractive alternatives but these reactions suffer from slow kinetics. Recently, Dawson and coworkers reported the dramatic acceleration of oxime formation in presence of aniline[13]; several applications of this reaction have subsequently been described.[14]

Here, we describe the synthesis of an aldehyde-containing analogue of FPP denoted as FAPP (**2**) and demonstrate that it is efficiently incorporated by PFTase. Model reactions with peptide substrates were used to confirm the structure of the prenylated product and monitor the rate of oxime formation. To investigate the utility of **2** for

protein modification, a GFP variant incorporating a CAAX-box was employed. PFTase was used to incorporate the modified isoprenoid into the protein. The resulting aldehyde-modified polypeptide was then either immobilized onto alkoxyamine-containing agarose beads or further functionalized with a second fluorophore via oxime formation.

Results and Discussion. Farnesyl aldehyde pyrophosphate (FAPP, **2**) was prepared in six steps starting from farnesol (Scheme 2.S1). Tetrahydropyran (THP)-protected farnesol was initially regioselectively oxidized at C-12 to a terminal alcohol followed by further oxidation to the corresponding aldehyde. Removal of the THP group, conversion to an allylic bromide and displacement with $[(n\text{Bu})_4\text{N}]_3\text{HP}_2\text{O}_7$ followed by purification by ion-exchange chromatography and reversed-phase HPLC (RP-HPLC) yielded FAPP **(2**) whose structure was confirmed by ^{1}H-NMR, ^{31}P-NMR, and HR-ESI-MS.

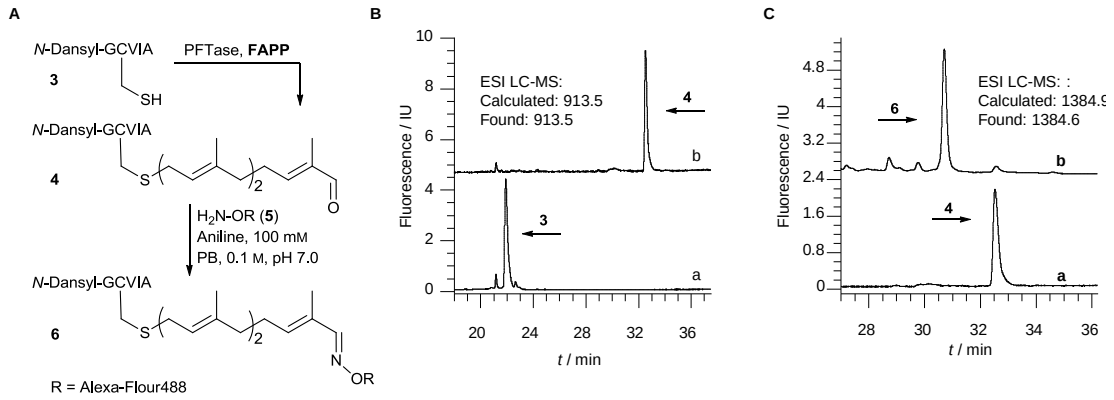


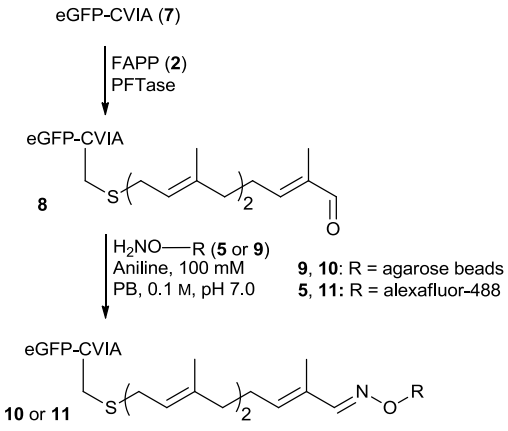
Figure 2.2. Prenylation of a CAAX-box containing peptide with FAPP (**2**) and subsequent aniline-catalyzed oxime ligation with aminooxy alexaflour-488. Panel A) Structures of reactants and products. Panel B) RP-HPLC analysis of the PFTase catalyzed prenylation of **3** with **2** to yield **4**. Chromatogram a: reaction mixture prior to the addition of enzyme; chromatogram b: reaction mixture after 40 min at 30 °C showing conversion to the prenylated peptide **4**. Panel C) RP-HPLC analysis of the oxime ligation reaction.

Reaction was initiated by the addition of aniline. Chromatogram a: reaction mixture prior to the addition of alexaflour-488 (**5**); chromatogram b: reaction mixture 40 sec after the addition of alexaflour-488 (**5**) and aniline to the reaction mixture at rt showing conversion to the oxime **6**. LC/MS analysis of the prenylated peptide and oxime ligated product confirmed quantitative formation of **4** and **6**.

Compound **2**, was first evaluated as a substrate for PFTase (Figure 2.2A) by employing a continuous fluorescence-based enzyme assay. Varying concentrations of **2** were incubated with a constant concentration of the fluorescent peptide substrate, *N*-dansyl-GCVIA (**3**) and PFTase (see Figure 2.S1 and 2.S2). From those experiments K_M and k_{cat} were found to be 2.2 μM and 0.065 μM min^{-1}, respectively. In comparison, similar experiments employing FPP (**1**) yielded values for K_M and k_{cat} of 0.42 μM and 0.26 μM min^{-1}, respectively. These values indicate a 5-fold increase in K_M, a 4-fold decrease in k_{cat} and an overall 20-fold drop in catalytic efficiency. While, significant, this attenuation can be limited to a 4-fold reduction in rate by performing preparative reactions at saturating substrate concentrations. To verify the results obtained from the kinetic analysis and confirm the structure of the prenylated peptide, the formation of peptide **4** was monitored by RP-HPLC analysis. The peptide substrate **3** (2.4 μM) was incubated with **2** (25 μM) and PFTase at 30 °C. Complete conversion was observed after 40 min (Figure 2.1B, chromatogram b). The reaction mixture was concentrated by solid phase extraction with a Sep-Pak cartridge and purified by RP-HPLC. LC-MS analysis of the purified material gave an $[M+H]^+$ of 913.5 consistent with the proposed structure of **4** (Figure 2.S3). Next the ligation reaction between aminooxy alexafluor **5** and aldehyde-functionalized peptide **4** was evaluated. Compound **4** (7.5 μM) was incubated with **5** (200 μM) in PB (0.1 M, pH

7.0) at rt and the reaction was initiated by addition of aniline (100 mM). To monitor the reaction, aliquots were withdrawn at 40 sec intervals and flash frozen. Subsequent RP-HPLC analysis showed that the ligation reaction was complete within the first 40 sec which is impressive for a coupling reaction at physiological pH in the µM concentration range of reagents (Figure 2.2C). LC-MS analysis of the reaction mixture gave an ion of 698.2 as the predominant species, which is consistent with $[M+2H]^{2+}$ for **6** (Figure 2.S3). Only a trace amount of unreacted starting aldehyde **4** was observed suggesting an equilibrium constant of 10^8 M for the ligation reaction. In contrast to the promising results noted above for oxime formation, attempts to achieve ligation via hydrazone formation between Texas-red hydrazine and **4** were not successful. RP-HPLC analysis of such a reaction mixture showed only a trace of possible product being formed. These results suggest that the equilibrium constant for hydrazone formation may not be sufficiently large to convert a significant percentage of aldehyde **4** when the reactants are present in µM concentrations.

With the ability of **2** to be incorporated by PFTase and subsequently derivatized via oxime formation established, we next evaluated the utility of this alternative substrate



Scheme 2.1. Prenylation of eGFP-CVIA with FAPP and subsequent immobilization on aminooxy functionalized agarose beads or labeling with aminooxy alexaflour-488 by oxime ligation.

for selective protein modification. Accordingly, **2** was incubated with eGFP-CVIA (**7**), in the presence of PFTase for 2 h at 30 °C (Scheme 2.1). The reaction time was based on our earlier observation that the peptide substrate **3** could be prenylated in less than 40 min. Concentration by ultracentrifugation followed by size exclusion chromatography to remove unreacted **2** yielded aldehyde-functionalized eGFP-CVIA (**8**).

To determine the efficiency of prenylation, thiol titration was performed on both unmodified protein **7** and the prenylated product **8**. DTNB titration of **7** revealed the presence of 2.8 thiols/protein in good agreement with the expected value of 3. A similar titration of **8** indicated the presence of 1.7 thiols/protein. These results indicate the loss of 1.1 thiols/protein after reaction and are consistent with complete modification of the cysteine residue present within the CVIA sequence after prenylation with **2**. Next, we examined two possible applications for the aldehyde-

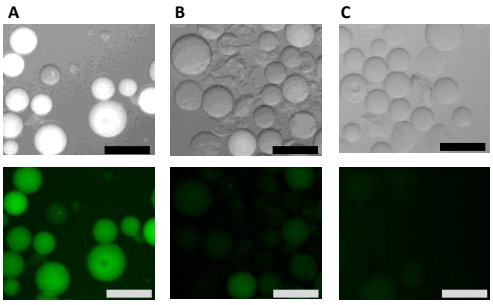


Figure 2.3. Immobilization of aldehyde-labeled eGFP-CVIA (**8**) onto aminooxy-functionalized agarose beads (**9**) to yield (**10**): Panel A) Reaction mixture in the presence of aniline (100 mM). Panel B) Reaction mixture in the absence of aniline. Panel C) Control immobilization reaction containing unmodified eGFP-CVIA (**7**) in the presence of aniline (100 mM). All reactions were carried out in the presence of protein (20 μM), and PB (100 mM, pH 7.0). Bright-field images are above and fluorescent microscope images below. Scale bars in the lower right-hand corners represent 200 μm.

modified protein produced above. First, its utility for protein immobilization was examined (Scheme 2.1). Aminooxy-functionalized agarose beads (**9**) were prepared by coupling a bifunctional aminooxy-PEG azide with alkyne-agarose beads via click chemistry. Beads (**9**) were incubated with aldehyde-functionalized eGFP-CVIA (**7**) at rt in the presence of 100 mM aniline. The reaction was vortexed and the fluorescence of supernatant was measured as a function of time. Results show that equilibrium is reached in less than 15 min. In that time, the beads became highly fluorescent (Figure 2.3A); less fluorescent beads were observed in the absence of aniline catalyst (Figure 2.3B) and no fluorescent beads were seen using GFP lacking the aldehyde moiety (Figure 2.3C). Based on the amount of **8** remaining in the supernatant, the efficiency of covalent immobilization (Figure 2.S5) was estimated to be ~30%.

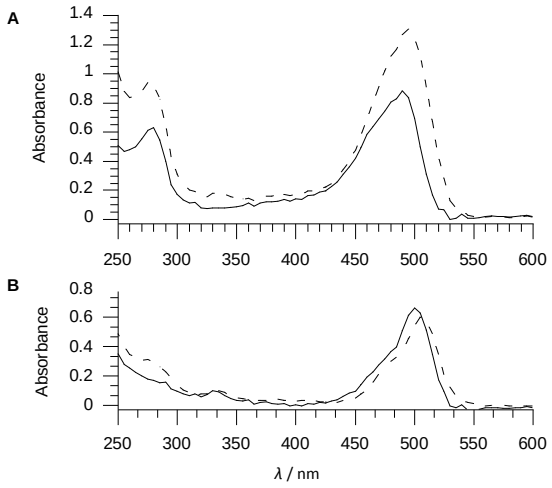


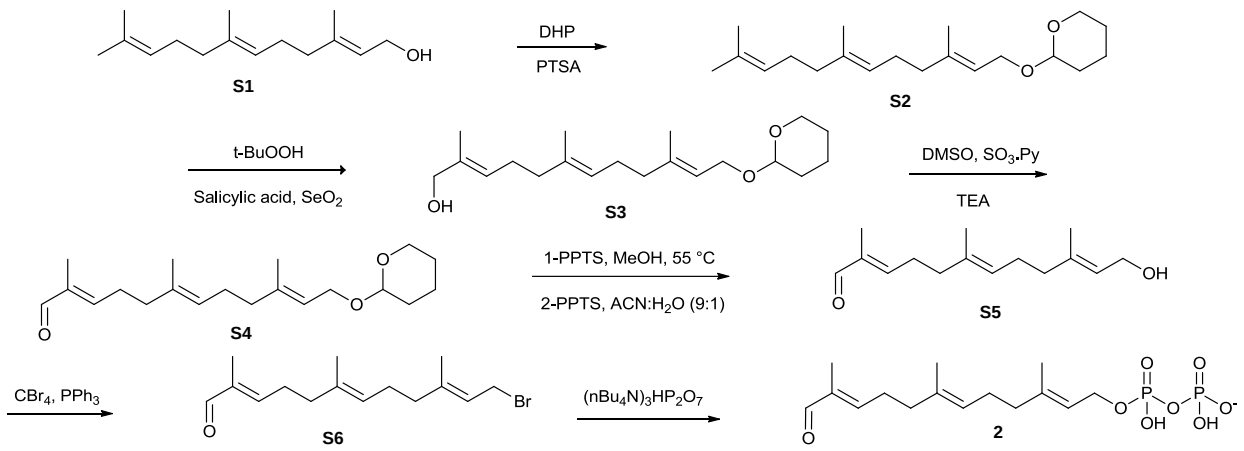
Figure 2.4. Selective labeling of aldehyde-labeled eGFP–CVIA (**8**) with alexaflour-488 (**5**) to yield **11**. Compound **8** (25 µM) was incubated with **5** (125 µM) and aniline (100 mM). After incubation at 25 °C for 30 min, samples were dialyzed against PB to remove unreacted **5**. Samples were then adjusted to contain equal protein concentration (17 µM) and the UV spectra were obtained. Panel A) Dashed line: Spectrum of **11** obtained from the treatment of **8** with **5**; Solid line: Spectrum of **8**. Panel B) Dashed line: Difference spectrum (Spectrum of **11** – Spectrum of **8**); Solid line: Spectrum of **5** (10 µM).

Finally, oxime ligation was used to site-specifically label eGFP-CVIA with a second fluorophore. Aminooxy alexafluor-488 (**5**) (125 μM) was treated with **8** (25 μM) in presence of 100 mM aniline for 30 min at rt. Labeled protein (**11**) was purified from excess **5** by extensive dialysis at 4 °C. The UV spectrum obtained for coupled product (**11**) showed a significant increase in absorbance near 500 nm when compared to the unmodified protein **8** (Figure 2.4A). The difference spectrum for this pair matches closely the spectrum of alexafluor-488 (**5**) thereby confirming successful ligation (Figure 2.4B). Quantitative comparison of the UV difference spectrum with the spectrum of alexafluor-488 suggests an efficiency of 60% for oxime ligation in this case.

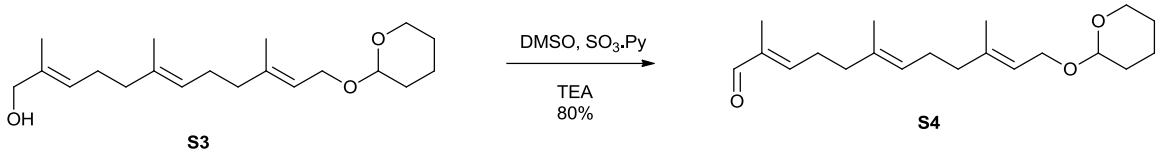
Conclusions. In this report, we demonstrate that PFTase can be used to introduce an aldehyde group near the C-terminus of a polypeptide and that the resulting protein can be immobilized or further functionalized via aniline-promoted oxime formation. Given that CAAX-box sequences can be appended to the C-termini of almost any protein, the method reported here should be useful for a wide range of applications in protein chemistry. This copper free labeling approach may be particularly useful for live cell studies or for applications where enzymatic activity must be preserved. It should also be noted that this labeling chemistry and the Cu(I)-catalyzed click reaction are orthogonal. This opens up the possibility of performing multiple modifications on proteins using different, bioorthogonal chemistries.

Supporting Information

General: All synthetic reactions were carried out at 25°C and stirred magnetically unless otherwise noted. TLC was performed on precoated (250 mm) silica gel 60 F-254 plates (Merck). Plates were visualized by staining with $KMnO_4$ or hand-held UV lamp. Flash chromatography silica gel (60–200 mesh, 75–250 μm) was obtained from Mallinckrodt Inc. CH_2Cl_2, CH_3CN, and THF were dried by using a Mbraun solvent purification system. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. ^1H NMR spectra were obtained at 300 or 500 MHz; ^{13}C NMR spectra were obtained at 125 MHz; ^{31}P NMR spectra were obtained at 121 MHz. All NMR spectra were acquired on Varian instruments at 25°C. Chemical shifts are reported in ppm and *J* values are in Hz. Fluorescence assay data were obtained by using a Varian Cary Eclipse Fluorescence Spectrophotometer. Analytical HPLC was performed on a Beckman model 125/166 instrument, equipped with a diode array UV detector, ABI Analytical Spectroflow 980 fluorescence detector, and a Varian C_{18} column (Microsorb-MV, 5 μm, 4.6x250 mm). Preparative HPLC separations were performed by using a Beckman model 127/166 instrument, equipped with a UV detector and a Phenomenex C18 column (Luna, 10 μm, 10x250 mm). MS spectra for synthetic reactions were obtained on a Bruker BioTOF II instrument. MS and LC/MS spectra of modified peptides were obtained with an Applied Biosystems/MDS SCIEX QSTAR® Elite Hybrid LC-MS system. Sep-Pak cartridges were purchased from Waters (Milford, MA). Yeast PFTase was prepared as previously described.[1]

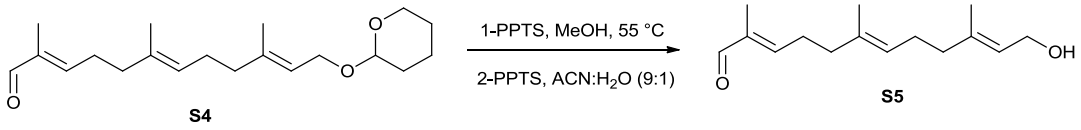


Scheme 2.S1. Syntheses of Farnesyl Aldehyde Pyrophosphate (FAPP, **2**).



(2E,6E,10E)-2,6,10-trimethyl-12-*O*-THP-dodeca-2,6,10-trienal (S4): Compound **S2** and **S3** were prepared as previously described.[2] Alcohol **S3** (1.00 g, 3.94 mmol) was dissolved in 20 mL anhydrous CH_2Cl_2 in a 50 mL flask and the solution was cooled to 0 °C in ice bath. DMSO (3.1 mL, 43 mmol) was added dropwise to the solution mixture followed by addition of 2.75 mL triethylamine (197 mmol). Sulfur trioxide pyridine complex (SO_3·Py, 2.52 g, 15.75 mmol) was added slowly over 10 min to the reaction mixture. The reaction was stirred at 0 °C for an additional h where TLC analysis showed almost complete conversion to the product. The reaction was stopped by addition of 100 mL CH_2Cl_2 and washed with 5 M HCl (2 x 10mL) until the aqueous phase remained acidic (via pH paper) which shows there is no more base left the in reaction mixture.

Next solution was washed with 15 mL $NaHCO_3$ followed by 2 x 10 mL brine. The solution was dried over Na_2SO_4 and solvent was removed *in vacuo*. Compound **S4** was purified by silica gel column chromatography using a step gradient of solvent (Hexane:EtOAc) starting from 1:0 (v/v) going to 3:1 (v/v) to afford 0.5 g of compound **S4** as a pale yellow oil (50% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 1.55 (m, 5H), 1.63 (s, 3H), 1.67 (s, 3H), 1.74 (s, 3H), 1.82 (m, 1H), 2.0-2.25 (m, 6H), 2.46 (m, 2H), 3.51 (m, 1H), 3.89 (m, 1H), 4.02 (dd, *J* = 7.3 Hz, *J* = 12 Hz, 1H), 4.24 (dd, *J* = 6.3 Hz, *J* = 12 Hz, 1H), 4.62 (dd, *J* = 3.0 Hz, *J* = 4.5 Hz, 1H), 5.16 (t, *J* = 6.2 Hz, 3 H), 5.36 (dd, *J*) 6.3 Hz, *J* = 7.3 Hz, 1H), 6.46 (t, *J*) 7.2 Hz, 1H), 9.37 (s, 1H). ^{13}C NMR (75 MHz, CDCl3) δ 195.38, 154.53, 139.95, 133.71, 125.33, 121.75, 120.86, 97.91, 63.70, 62.35, 39.51, 38.01, 30.77, 27.45, 26.25, 25.55, 19.69, 16.47, 15.97, 9.28. HR-ESI-MS calcd for $C_{20}H_{32}O_3Na$ $[M+Na]^+$ 343.2249, found 343.2246.

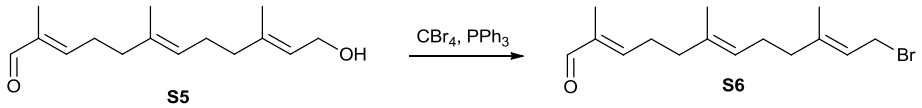


S4 **S5**

(2E,6E,10E)-12-hydroxy-2,6,10-trimethyldodeca-2,6,10-trienal (S5): Protected aldehyde **S4** (0.66 g, 2.6 mmol) was dissolved in 15 mL CH_3OH in a 25 mL flask. PPTS (40 mg) was added as catalyst. The reaction was then refluxed at 55 °C for 3 h. It was then quenched by adding 10 mL sat. $NaHCO_3$ and 100 mL EtOAc. The organic layer was separated and dried over Na_2SO_4. Early ^{1}H NMR analysis revealed deprotection of THP group in high yield but also protection of the aldehyde to a acetal with two methoxy groups as was expected. The product was dissolved in 15 mL of CH_3CN and 1.5 mL of

H_2O. PPTS (25 mg) was then added and the reaction was stirred overnight at rt. 100 mL EtOAc was added to the reaction which was then washed with 10 mL sat. $NaHCO_3$ followed by 2 x 10 mL brine. The organic phase was collected and dried over Na_2SO_4. Solvent was evaporated *in vacuo* and the product was further purified by silica gel column chromatography with gradient elution (hexane:EtOAc) from 10:1 (v/v) going to 2:1 (v/v) to afford 0.22 g of compound **S5** as a pale yellow oil (38% yield). It should be mentioned that aldehyde deprotection was only 50% achieved when THF was used instead of CH_3CN as solvent.

^{1}H NMR (500 MHz, CDCl3) δ 1.64 (s, 3H), 1.68 (s, 3H), 1.74 (s, 3H), 2.0-2.25 (m, 6H), 2.45 (t, J = 7.5, 2H), 4.143 (t, J = 7 Hz, 1H), 5.16 (t, J = 7 Hz, 3 H), 5.41 (t, J = 4.5), 6.48 (t, J = 7.5 Hz, 1H), 9.37 (s, 1H). ^{13}C-NMR (75 MHz, CDCl3) δ 195.352, 154.618, 139.262, 133.602, 125.160, 123.718, 117.303, 63.569, 39.334, 38.084, 27.336, 26.174, 16.203, 15.851, 9.163. HR-ESI-MS calcd for $C_{15}H_{24}O_2Na$ $[M+Na]^+$ 259.1674, found 259.1669.

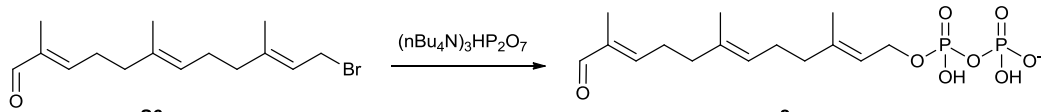


S5 **S6**

(2E,6E,10E)-12-bromo-2,6,10-trimethyldodeca-2,6,10-trienal (S6): PPh_3 (polymer-supported beads, 167 mg, 0.5 mmol) was added to a solution of **S5** (107 mg, 0.454 mmol) in CH_2Cl_2 (15 mL), and the reaction was stirred for 20 min at room temperature to let the beads swell in the solution. CBr_4 (165 mg, 0.5 mmol) was added slowly to the reaction mixture and the reaction was stirred for 15 min when TLC analysis showed most of starting material was converted to product. The reaction mixture was filtered to separate

beads, and solvent was evaporated *in vacuo*. The product was further purified by silica gel flash chromatography to yield 120 mg of **S6** as a pale yellow liquid. (86%).

^{1}H NMR (500 MHz, CDCl3) δ 1.628 (s, 3H), 1.723 (s, 3H), 1.738 (s, 3H), 2.0-2.25 (m, 6H), 2.45 (t, J = 7.5, 2H), 4.013 (t, J = 8.5 Hz, 2H), 5.125 (t, J = 7 Hz, 1 H), 5.520 (t, J = 4.5, 1H), 6.462 (t, J = 7.5 Hz, 1H), 9.382 (s, 1H). ^{13}C-NMR (75 MHz, CDCl3) δ 195.288, 154.450, 141.242, 134.066, 124.679, 121.644, 120.763, 39.329, 38.076, 37.940, 27.392, 26.951, 15.995, 15.947, 9.259.



(2E,6E,10E)-3,7,11-trimethyl-12-oxododeca-2,6,10-trien-1-yl dihydrogen diphosphate (2):

Compound **S6** (100 mg, 0.317 mmol) and $[(n\text{Bu})_4\text{N})]_3\text{HP}_2\text{O}_7$ (902 mg, 1.0 mmol) were mixed in dry CH_3CN (1 mL) and stirred at rt for 3 h, after which the solvent was removed *in vacuo*. Bio-Rad AG 50W-X8 ion-exchange resin (100–200 mesh, H^+ form) was used to convert the product to its ammonium form. The resin was packed and washed with three column volumes of H_2O/conc. NH_4OH (2:1, v/v) until the elution solvent became basic. The resin was then equilibrated with four volumes of NH_4HCO_3 (25 mm)/*i*PrOH (49:1, *v/v*; solvent A). The crude product was dissolved in 1 mL of solvent A and applied to the column and eluted with additional solvent A (20 mL), lyophilized, and purified by RP-HPLC with a semipreparative column using the following conditions: detection: 214 nm; flow rate: 5.0 mL min^{-1}; 2 mL injection loop; gradient 0–30% solvent B in 30 min, 60–100% in 5 min; solvent A: 25 mm NH_4HCO_3, solvent B: CH_3CN. Compound **2** eluted

from 20–25% solvent B. Fractions containing pure **2** were collected and lyophilized to yield 0.7 mg (0.54%) of a white powder.

^{1}H NMR: (500 MHz, D_2O) δ 1.488 (s, 3H), 1.545 (s, 6H), 1.917 (t, J = 7 Hz, 2H), 1.996 (t, J = 7 Hz, 2H), 2.067 (t, J = 7 Hz, 2H), 2.377 (t, J = 7 Hz, 2H), 4.302 (t, J = 6 Hz, 2H), 5.081 (t, J = 6 Hz, 1H), 5.293 (t, J = 7 Hz, 1H), 6.613 (t, J = 7.5 Hz, 1H), 9.108 (s, 1H). ^{31}P NMR: (121 MHz, D_2O) δ -5.971 (d, J = 22.6, 1P), -10.013 (d, J = 22.6, 1P). HR-ESI-MS calcd for $C_{15}H_{26}O_8P_2$ $[M-H]^-$ 395.1025, found 395.0907.

Enzymatic studies of FAPP (2) using a continuous fluorescence assay: Enzymatic reaction mixtures contained 50 mM Tris·HCl, pH 7.5, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 5.0 mM DTT, 2.4 μM *N*-dansylGCVIA, 0.040 % (*w/v*) *n*-dodecyl-ß-D-maltoside, 80 nM PFTase, and varying concentrations of **2** (0-10 μM), in a final volume of 450 μL. The reaction mixtures were equilibrated at 30 °C for 5 min, initiated by the addition of PFTase, and monitored for an increase in fluorescence (λ_{ex} = 340 nm, λ_{em} = 505 nm, slit widths = 10 nm for both) for approximately 10 min. The initial rates of formation of products were obtained as slopes in IU/min using least squares analysis. Correction was applied to all rate calculation based on the difference of intensity of fluorescence of product and starting peptide. The difference corresponds only to the fluorescence of total amount of the product. The slope was then divided to the difference followed by multiplying to the total concentration of peptide which then gives the relative rate of formation of product. It should be noted that the K_M values reported here are actually apparent K_M values since the measurements were performed at only a single peptide

concentration. The data were fit to a Michaelis-Menten model using KaleidaGraph, a nonlinear regression program, to determine K_M.

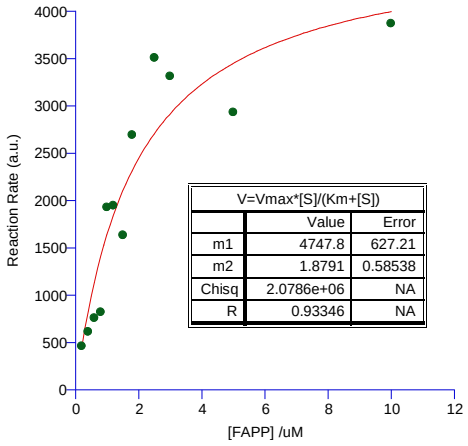


Figure 2.S1. Fluorescence-based PFTase enzyme assay for prenylation of **3** using varying concentrations of **2**.

To compare the catalytic efficiency of incorporation of FPP and **2**, enzymatic reactions were performed as described above with the exception of the isoprenoid substrates, FPP or **2**, which were included at 10 or 30 µM, respectively. Reaction mixtures were monitored spectrofluorometrically for 1 h to insure the endpoint was reached. The rate of reaction was then determined from the initial slope.

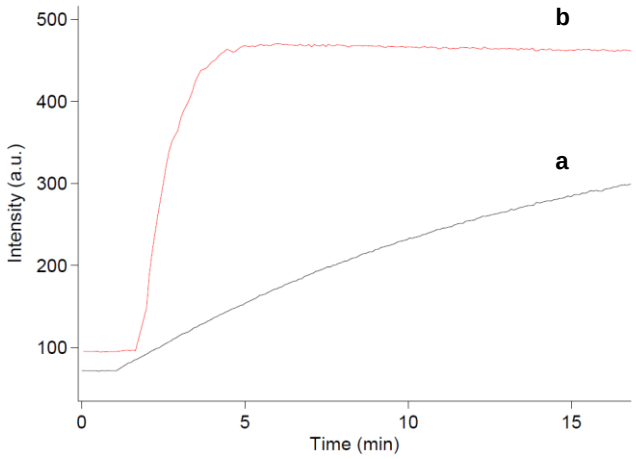
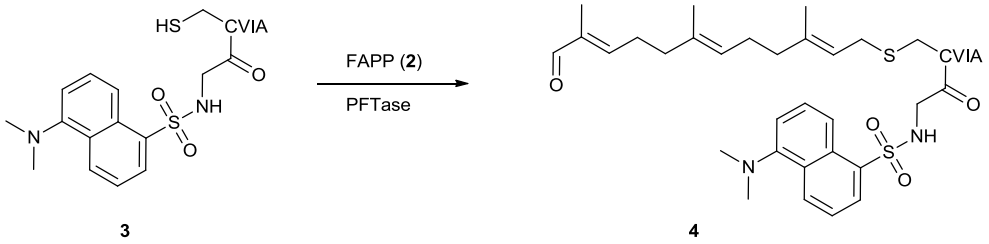


Figure 2.S2 Continuous fluorescence assay for PFTase catalyzed peptide prenylation of peptide 3 with 4 μM FPP or 10 μM FAPP at same concentration of enzyme (80 nM). **a**: Reaction of **2** (30 μM). **b**: Reaction of FPP (10 μM).



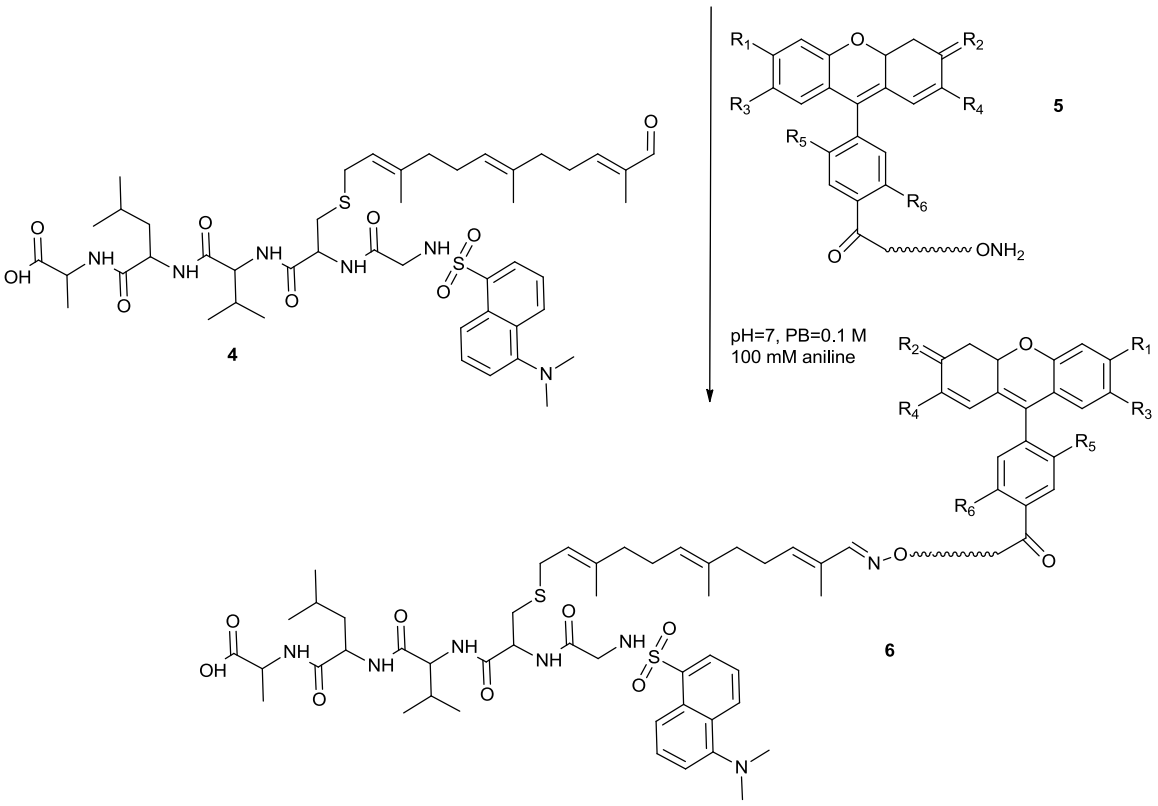
3 4

Enzymatic synthesis of 4: Enzymatic reactions (26 mL) contained Tris-HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), $ZnCl_2$ (10 mM), DTT (5.0 mM), **3** (2.4 μM), PFTase (80 nM), and **2** (30 μM). To ensure complete disulfide reduction of the peptide, all reagents except FAPP and enzyme were premixed and incubated, for 2 h at 4 °C. With all reagents mixed, the reaction was initiated by the addition of enzyme and the resulting mixture was incubated at 30°C for 1 h. The reaction progress was monitored by fluorescence detection (λ_{ex}=340 nm, λ_{em}=504 nm) using analytical RP-HPLC. The following conditions were employed: flow rate: 1 mL min^{-1}; 100 mL injection loop; gradient 0–100% B in 40 min;

solvent A: NH_4HCO_3 25 mM in H2O; solvent B: CH3CN. After 1 h, the reaction was purified by using a Waters Sep-Pak Plus reversed-phase C_{18} Environmental Cartridge. The cartridge was first washed with solvent B (10 mL) followed by equilibration with solvent A (20 mL). The crude enzymatic reaction mixture was applied to the cartridge and a gradient elution was performed in the following sequence: 10 mL solvent A, 10 mL solvent C (20% solvent B, 80% solvent A), 10 mL solvent D (40% solvent B, 60% solvent A), 10 mL solvent E (60% solvent B, 40% solvent A). Fractions (1 mL) were collected and product elution was monitored by using a handheld UV lamp. The green-fluorescent product was clearly visible and the brightest fraction was selected and its purity was confirmed by HPLC.

LC-MS analysis of the purified product gave an ion of 913.5 as the predominant species, which is consistent with $[M+H]^+$ for **4** (Figure 2.S3A).

Oxime ligation between peptide-aldehyde 4 and aminooxy alexafluor-488 (5):



Coupling reactions contained 7.5 µM **4**, 200 µM alexafluor-488 (**5**), 0.1 M PB, pH 7.0, and 100 mM aniline in final volume of 500 µL. Reactions were performed at rt and were initiated by addition of aniline (100 mM). To monitor the reactions, aliquots were withdrawn at 40 sec intervals and flash frozen by liquid nitrogen. LC-MS analysis of the reaction mixture gave an ion of 698.2 as the predominant species, which is consistent with $[M+2H]^{2+}$ for **6** (Figure 2.S3B).

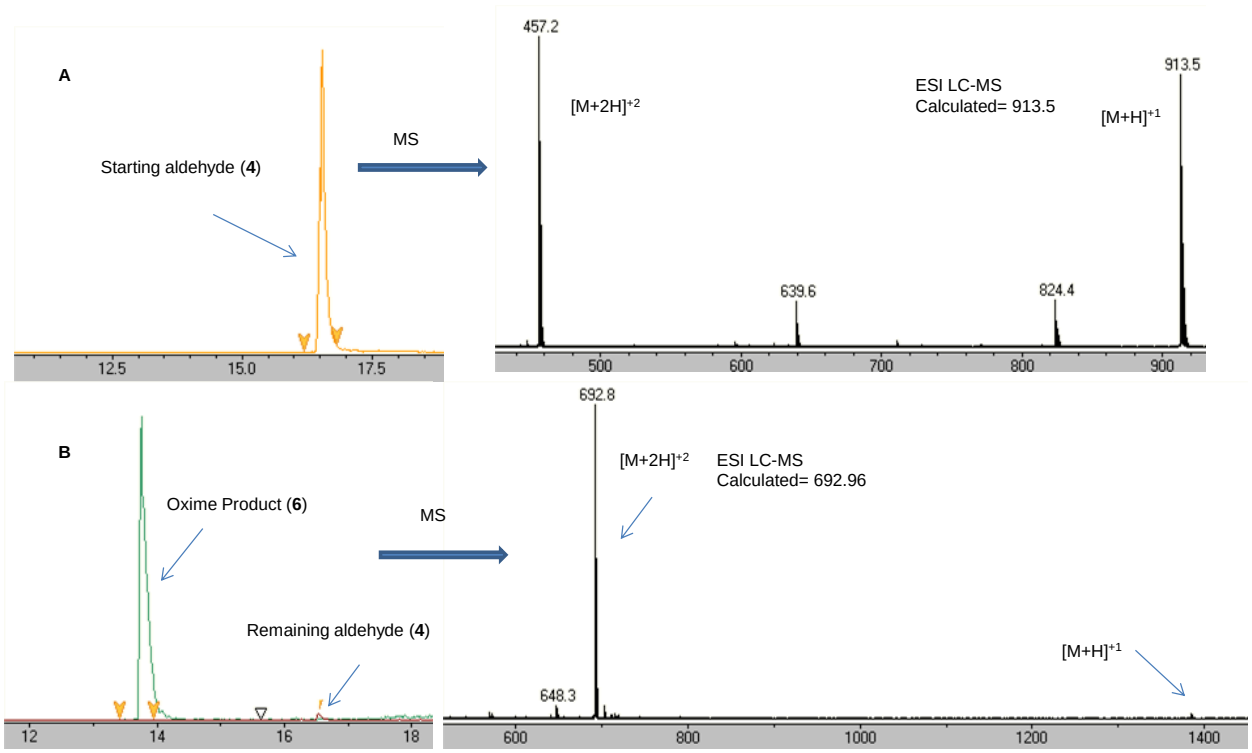


Figure 2.S3 ESI/LC-MS analysis of oxime ligation reaction between peptide aldehyde **4**, and alexaflour-488 (**5**). Left panels are extracted ion chromatograms and right panels are corresponding MS spectra. **A**) Extracted ion chromatogram of starting peptide-aldehyde **4** and corresponding MS spectrum. **B**) Extracted ion chromatogram of coupled product (**6**) superimposed on extracted ion chromatogram of starting aldehyde (**4**). Both chromatograms are extracted from the total ion chromatogram of the reaction mixture after 1 h of reaction. Relative intensity of peaks shows completion of oxime ligation to 99% conversion. Corresponding MS spectrum of the oxime product (**6**) confirms the formation of oxime bond.

eGFP-CVIA: Protein was prepared as previously described with one modification.[3] In the final phenyl sepharose chromatography step, after the protein was loaded onto the column and washed with buffer as explained in the original work, the protein was eluted from column by adding water (no buffer). Protein could not be eluted from the column by adding buffer only (as it was described in the original reference).

Enzymatic incorporation of FAPP moiety into eGFP-CVIA: Enzymatic reactions (10 mL) contained Tris-HCl (50 mM, pH 7.0), $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), eGFP–CVIA (2.0 μM), **2** (30 μM), and PFTase (80 nM). After incubation at 30°C for 3 h, the reaction mixture was concentrated using an Amicon Centriprep centrifugation device (10,000 MW cutoff), and excess **2** was removed by using a NAP-5 column (Amersham). The subsequent protein concentration was calculated based on volume change assuming 100% recovery.

Thiol titration of eGFP for determining efficiency of prenylation: $KHPO_3$ (0.1 M, pH 7.4, KPB) was used as buffer; 5 mM DTNB solution was prepared by dissolving 99 mg of DTNB in 50 mL of KPB. Dissolution of DTNB is slow and requires vortexing for a few minutes to become completely clear. Protein free of exogenous thiols (15 μL, prepared by NAP-5 desalting) was added to a solution of 75 μL of 6 M Guanidine•HCl and 10 μL of above DTNB solution. The final mixture was vortexed and the absorbance at 412 nm was measured after 5 min. Each measurement was repeated 3 times. A standard curve was prepared using cysteine as a standard. All measurements were repeated in triplicate.

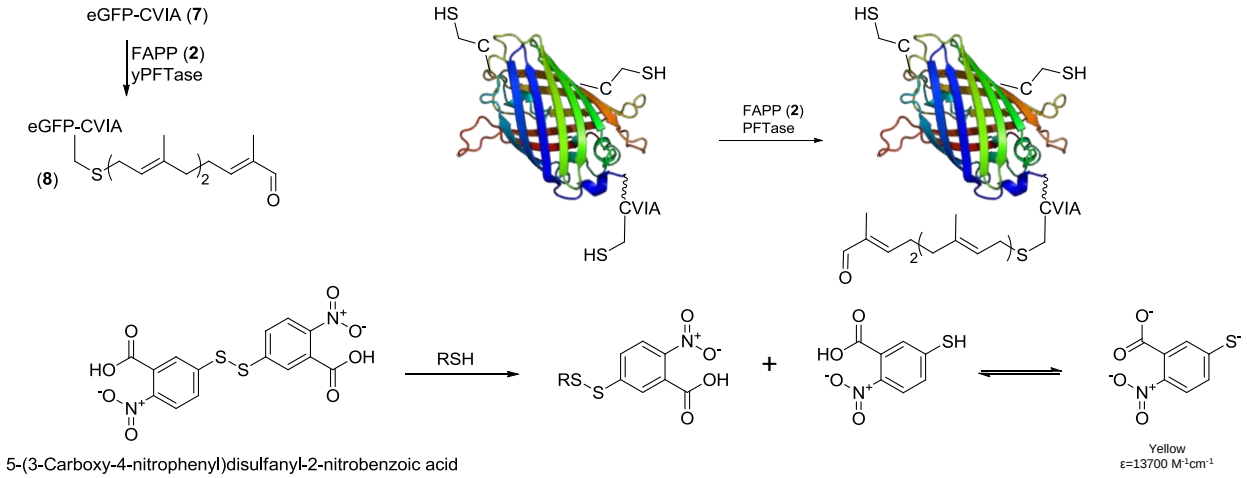
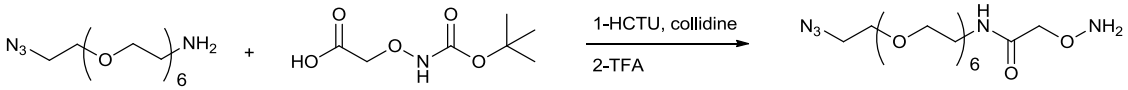


Figure 2.S4. Thiol titration of prenylated and unprenylated eGFP-CVIA by DTNB.

Aminooxy agarose beads (11): Alkyne beads were prepared as previously described.[3] Aminooxy beads were prepared by treating aminooxy-PEG azide (**S9**) with alkyne agarose resin under click chemistry conditions. Aminooxy-PEG azide was synthesized from amine-PEG azide (**S7**) by coupling to N-Boc-aminooxy-acetic acid (**S8**) followed by Boc removal with TFA.



S7 **S8** **S9**

PEG **S7** (100 mg, 0.29 mmol) was added to a solution of 360 mg HCTU (0.87 mmol), 268 µL of collidine[4] (2.03 mmol) and 166 mg **S8** (0.87 mmol) in 1 mL of DMF. The reaction was allowed to proceed for 2 h. TFA (3 mL) dissolved in CH_2Cl_2 (4 mL) was added to mixture and the reaction was allowed to proceed for an additional h. Solvent was removed *in vacuo* to give **S9** which was used without further purification. ^{1}H NMR

(500 MHz, $CDCl_3$) δ 3.177 (m, 4H), 3.381 (b, 24H), 4.425 (s, 2H), 8.090 (s, 1H). HR-ESI-MS calcd for $C_{16}H_{34}O_8N_5$ $[M+H]^+$ 424.2407, found 424.2435. ^1H NMR and MS analysis show ~complete conversion of starting amino-PEG **S7** to final aminooxy-PEG **S9**.

S10 **9**

One tenth of the product solution of **S9** (0.03 mmol PEG) was removed and the pH was adjusted with a mixture of PB and NaOH to neutrality. Alkyne beads **S10** (1 mL) in PB, 0.1 M, pH 7 was added to this solution. $CuSO_4$ (1 mM), TCEP (1 mM), TBTA (100 µM) were added and the reaction was allowed to proceed for 2 h. Beads were centrifuged down and the overlaying solution was exchanged for PB. The beads were centrifuged and the buffer exchanged again. This was repeated 5 times. Beads were stored in PB solution at 4 °C for future use.

Immobilization of eGFP onto aminooxy agarose beads: To immobilize the protein onto the agarose beads, aldehyde-functionalized eGFP–CVIA (**8**, 50 µL, 40 µM stock in PB) was added to beads (**9**) in PB (20 µL). The ligation reaction was initiated by the addition of 100 mM aniline.

For control reactions, eGFP-CVIA (**7**) was added instead. GFP fluorescence of the supernatant was measured as a function of time. Beads were centrifuged and 5 µL of

supernatant was removed, diluted into 445 µL of PB and the fluorescence was measured (λ_{ex}=488 nm, λ_{em}=510 nm).

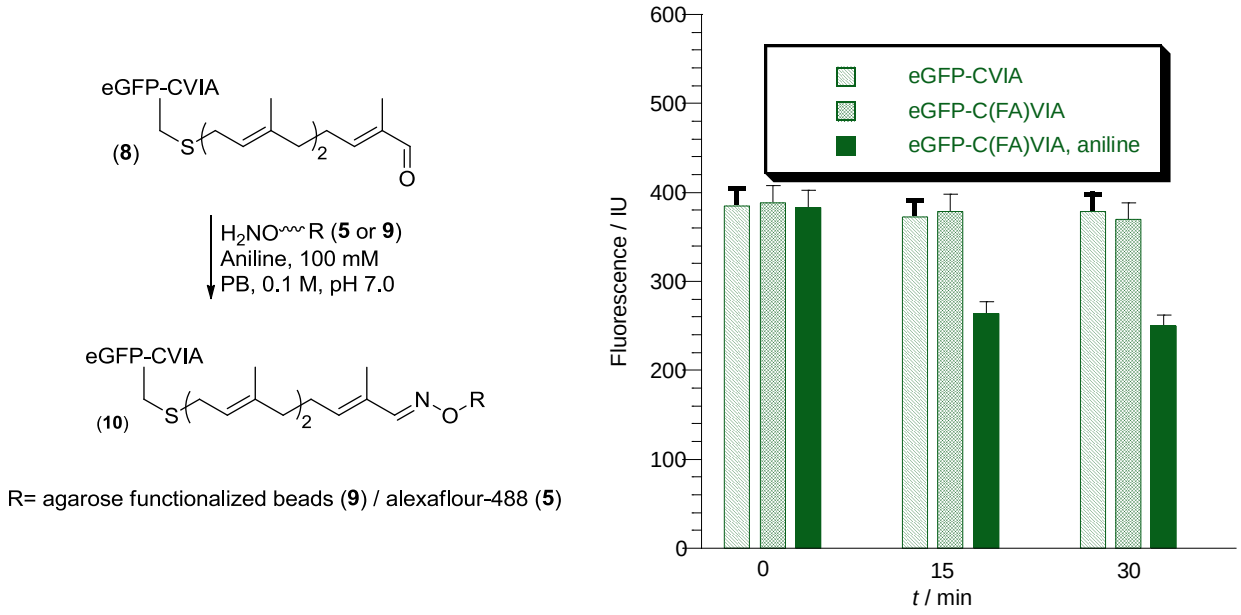


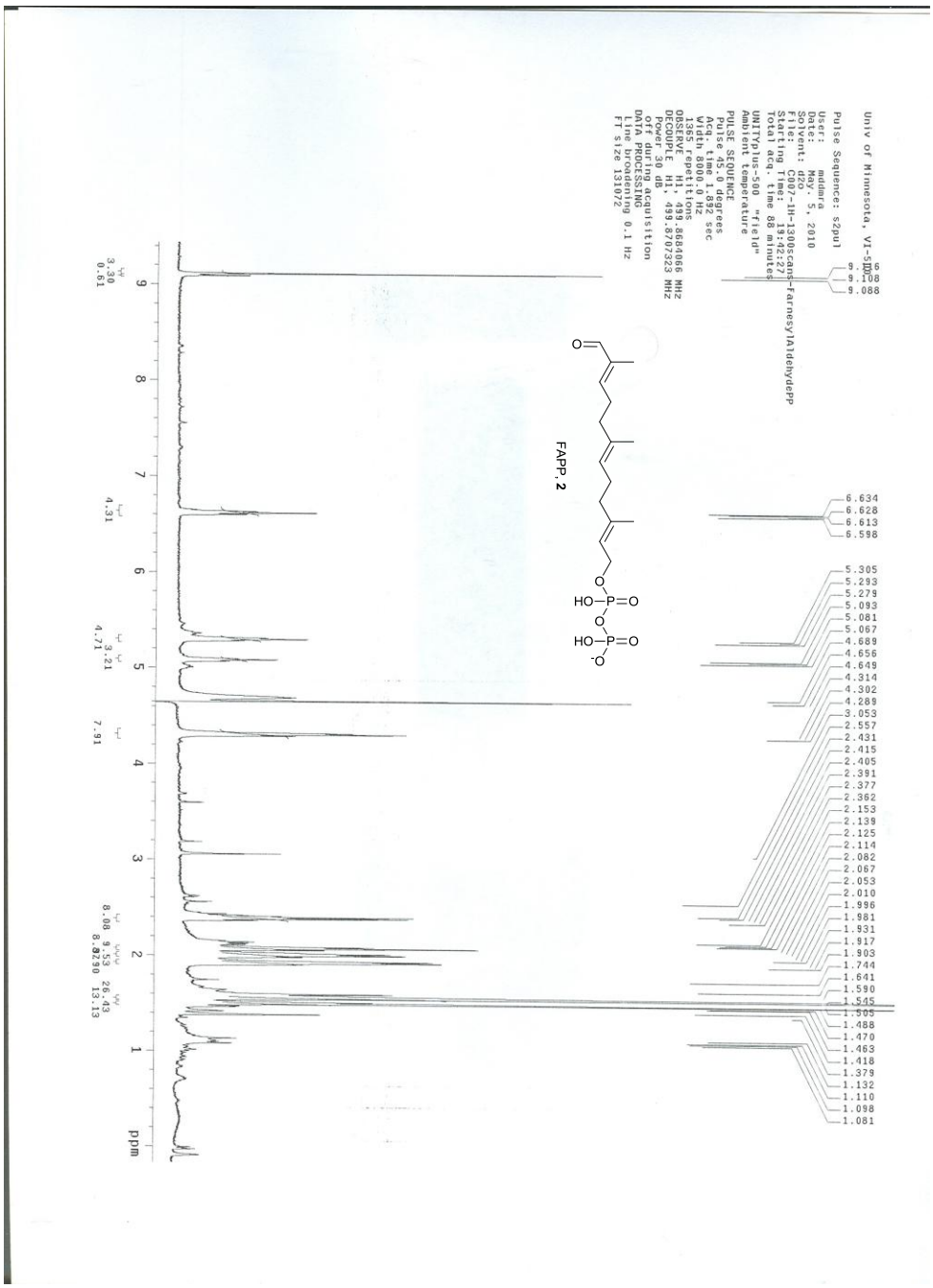
Figure 2.S5. Immobilization efficiencies of aldehyde-labeled eGFP **8** onto aminooxy agarose beads (**9**). To determine the immobilization efficiencies and background labeling, eGFP-CVIA and aldehyde functionalized eGFP–CVIA (30 µM) were incubated with or without 100 mM aniline. At the appropriate times, beads were centrifuged and the fluorescence remaining in the supernatant was determined by spectrofluorimetry.

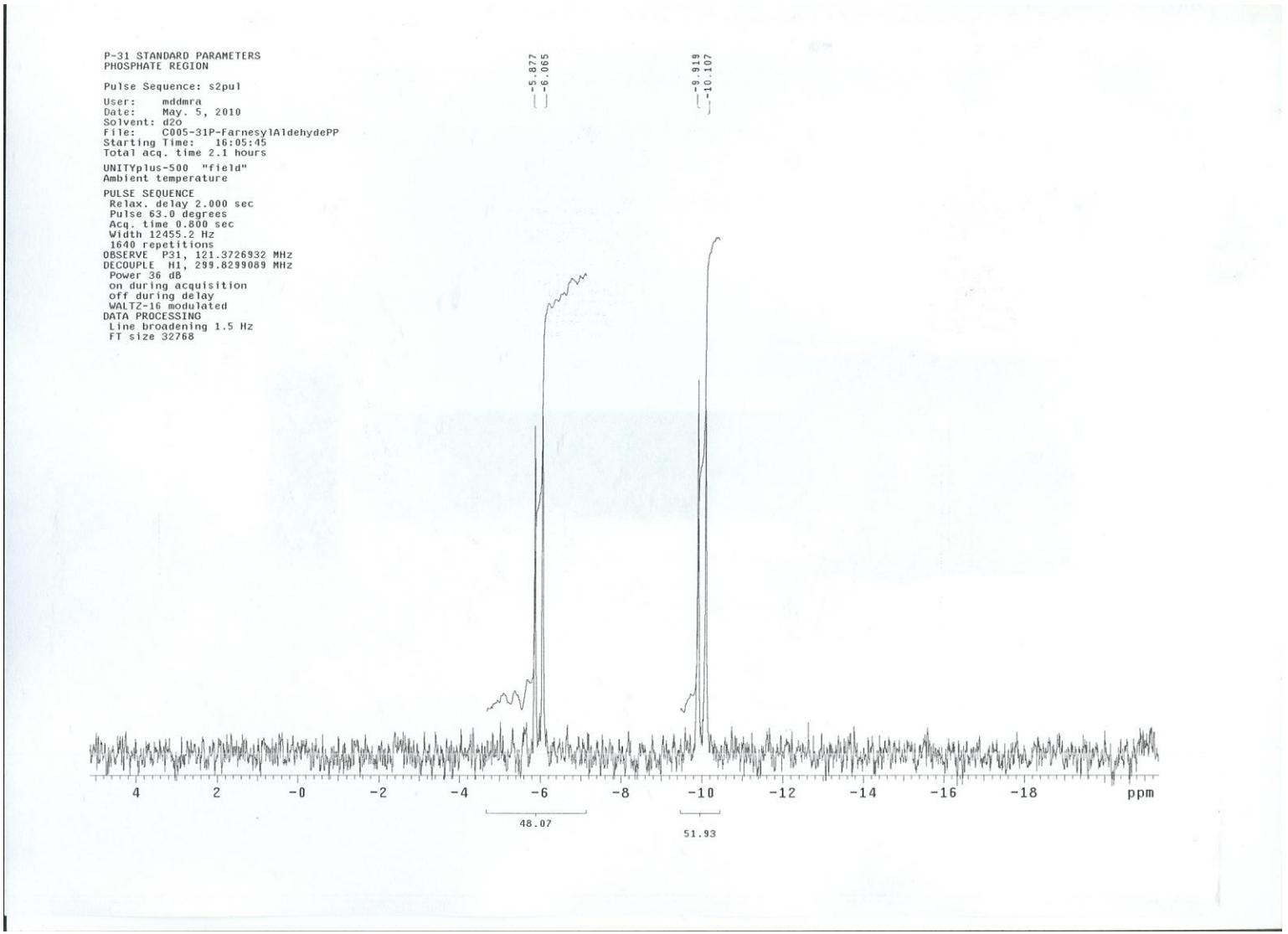
Coupling reaction between aldehyde-labeled eGFP-CVIA and alexafluor-488:

Alexafluor-488 (**5**) (7 µL of 1.9 mM solution in DMSO was added to 100 µL of eGFP-C(FA)VIA (**8**) (stock solution of ~30 µM in PB). The reaction was initiated by adding 100 mM aniline and was allowed to proceed for 30 min at rt. The mixture was then transferred to a micro dialysis cassette (10,000 MWCO, Pierce) and dialyzed against PB (15 mL) at 4°C for 3 days. The control reaction contained 7 µL of DMSO instead of the **5**. After dialysis, the final volume of sample was measured to be 165 µL. The control

reaction solution was also diluted to 165 µL and UV spectra were obtained. The concentrations of eGFP in both samples were determined by Bradford assay to be 17.1 µM for the control and 17.9 µM for the coupling reaction.

NMRs




P-31 STANDARD PARAMETERS
PHOSPHATE REGION
Pulse Sequence: s2pul
User: mddmra
Date: May. 5, 2010
Solvent: d2o
File: C005-31P-FarnesylAldehydePP
Starting Time: 16:05:45
Total acq. time 2.1 hours
UNITYplus-500 "field"
Ambient temperature
PULSE SEQUENCE
Relax. delay 2.000 sec
Pulse 63.0 degrees
Acq. time 0.800 sec
Width 12455.2 Hz
1640 repetitions
OBSERVE P31, 121.3726932 MHz
DECOUPLE H1, 299.8299089 MHz
Power 36 dB
on during acquisition
off during delay
WALTZ-16 modulated
DATA PROCESSING
Line broadening 1.5 Hz
FT size 32768
-5.877
-6.065
-9.919
-10.107
4
2
-0
-2
-4
-6
-8
-10
-12
-14
-16
-18
ppm
48.07
51.93

Chapter 3. Chemoenzymatic Reversible Immobilization and Labeling of Proteins without Prior Purification

Mohammad Rashidian, Mark D. Distefano

Site-specific chemical modification of proteins is important for many applications in biology and biotechnology. Recently, our laboratory and others have exploited the high specificity of the enzyme protein farnesyltransferase (PFTase) to site-specifically modify proteins through the use of alternative substrates that incorporate bioorthogonal functionality including azides and alkynes. In this study, we evaluate two aldehyde-containing molecules as substrates for PFTase and as reactants in both oxime and hydrazone formation. Using green fluorescent protein (GFP) as a model system, we demonstrate that the purified protein can be enzymatically modified with either analogue to yield aldehyde-functionalized proteins. Oxime or hydrazone formation was then employed to immobilize, fluorescently label or PEGylate the resulting aldehyde-containing proteins. Immobilization via hydrazone formation was also shown to be reversible via transoximization with a fluorescent alkoxyamine. After characterizing this labeling strategy using pure protein, the specificity of the enzymatic process was used to selectively label GFP present in crude *E. coli* extract followed by capture of the aldehyde-modified protein using hydrazide-agarose. Subsequent incubation of the immobilized protein using a fluorescently labeled or PEGylated alkoxyamine resulted in the release of pure GFP containing the desired site-specific covalent modifications. This procedure was also employed to produce PEGylated glucose-dependent insulinotropic polypeptide (GIP), a protein with potential therapeutic activity for diabetes. Given the specificity of the PFTase-catalyzed reaction coupled with the ability to introduce a CAAX-box recognition sequence onto almost any protein, this method shows great potential as a general approach for the selective immobilization and labeling of

recombinant proteins present in crude cellular extract without prior purification. Beyond generating site-specifically modified proteins, this approach for polypeptide modification could be particularly useful for large scale production of protein conjugates for therapeutic or industrial applications.

Introduction.

Site-specific chemical modification of proteins is important for many applications in biology and biotechnology. It can facilitate studies of proteins with respect to their structure, folding, and interaction with other proteins in both biochemical and cellular investigations. In particular, in many biotechnology applications, the oriented (i.e. site-specific) covalent attachment of proteins to surfaces is important because it ensures homogeneous surface coverage and accessibility to the active site of the protein.[11,109–115] Protein immobilization is an important first step for many applications including the construction of biosensors and protein microarrays, development of immunoassay methods, and employment of enzymes in biotechnology procedures.[13,116,117] Similarly, site-specific protein labeling is essential for a variety of applications ranging from the introduction of fluorophores for biophysical studies to the preparation of protein-polymer conjugates for medical applications.[49,118–122]

Importantly, the structural sensitivity of polypeptides calls for chemical transformations that proceed under mild conditions and that are compatible with all functional groups present therein. However, such modification is challenging because of the large number of reactive functional groups typically present in polypeptides. Although many existing chemical reactions are applicable in principle, the development of new methods for site-

specific modification of proteins that function under mild conditions is an area of intense research.[123] While a number of reactions suitable for protein modification have been developed,[29,31,124,125] to date, the Cu(I) catalyzed click reaction has been the most widely employed bioorthogonal process.[29] Although highly useful, that reaction employs Cu(I) which is toxic to cells and can in some cases erode biological activity. To address those issues, copper-free variations of the click reaction have been developed that function based on the inclusion of electron withdrawing substituents and/or ring strain into alkyne-containing reagents. While highly promising, these new reagents are not generally commercially available, are difficult to synthesize and manifest low aqueous solubility.[126–128]

As an alternative, oxime and hydrazone-based reactions have found wide application in the conjugation of biomolecules on account of the absence of aldehyde or ketone groups in proteins and their orthogonal reactivity with aminooxy or hydrazine derivatives to give stable hydrazones or oximes.[26–33] While reactions between aldehydes and ketones with alkoxyamines or hydrazides are generally slow, they can be significantly accelerated by the addition of aniline.[38,129] This has resulted in a number of exciting applications ranging from the site-specific glycosylation of proteins[130] to the fluorescent labeling of bacteria and mammalian cells.[131] Given the utility of oxime and hydrazone formation, a number of methods have been developed to introduce aldehydes and ketones into proteins. Chemical approaches include transamination in the presence of sodium glyoxylate and copper sulfate[132] or using pyridoxal-5-phosphate;[133] while these have proved to be powerful methods, they are not applicable to all *N*-termini and cannot always be driven to

completion. Enzymatic methods include the action of formylglycine-generating enzyme[44] or nonsense suppression approaches that permit the incorporation of aldehydes into internal positions within proteins.[134]

Recently, our laboratory and others have exploited the high specificity of the enzyme protein farnesyltransferase[135] (PFTase) to site-specifically modify peptides and proteins.[32,41–43] In nature, PFTase catalyzes the transfer of a farnesyl isoprenoid group from farnesyl diphosphate (FPP, Figure 3.1) to a sulfur atom present in a cysteine residue. That residue must be located in a tetrapeptide sequence (denoted as a CAAX-box) positioned at the C-terminus of a protein or peptide to be a PFTase substrate. Interestingly, CAAX-box sequences such as CVIA can be appended to the C-termini of many proteins rendering them efficient substrates for PFTase. Since PFTase can tolerate many simple modifications to the isoprenoid substrate,[3,44–48] it can be used to introduce a diverse range of functionality into proteins at their C-termini. Chemoselective reaction with the resulting functionalized protein can then be used for a wide range of applications. Since the "AAX" residues from a CAAX-box sequence can be removed by treatment with carboxypetidase after prenylation, the net addition to the protein in this labeling method can be limited to a single prenylcysteine residue.[93]

In an initial communication,[82] we reported that compound **1**, an aldehyde-containing analogue of FPP, can be incorporated into a purified protein substrate using PFTase and that the resulting aldehyde-functionalized protein can be immobilized or fluorescently labeled via oxime formation. In this study, we have followed up on those initial observations by comparing the properties of α,β-unsaturated aldehyde **1** with aryl

aldehyde **2** in terms of their efficiency as PFTase substrates and as reactants in both oxime and hydrazone formation. Using green fluorescent protein (GFP[136]) as a model system, we demonstrate that the purified protein can be enzymatically modified with **1** or **2**. Oxime or hydrazone formation was then employed to immobilize, fluorescently label or PEGylate the resulting aldehyde-functionalized proteins. Immobilization via hydrazone formation was also shown to be reversible via transoximization with a

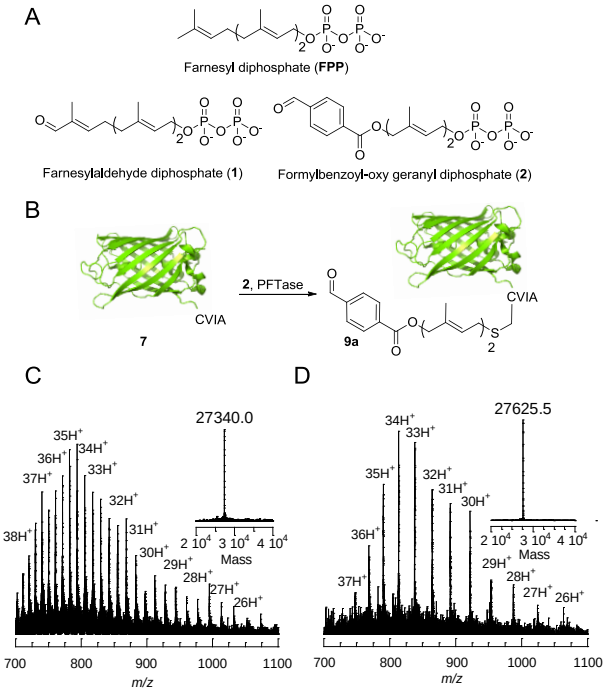


Figure 3.1. A) Structures of farnesyl diphosphate, farnesyl aldehyde diphosphate (**1**) and formylbenzoyl-oxy geranyl diphosphate (**2**). B) Schematic representation of prenylation of a protein contaning a CAAX-box positioned at its C-terminus (GFP-CVIA, **7**) with aldehyde-containing analogue **2** to yield the prenylated product **9a**. C) ESI MS analysis of **7** with the deconvoluted mass spectrum shown in the inset. D) ESI MS analysis of **9a** with the deconvoluted mass spectrum shown in the inset.

fluorescent alkoxyamine. After characterizing this labeling strategy using pure protein, the specificity of the enzymatic process was used to selectively label GFP present in crude *E. coli* extract followed by capture of the aldehyde-modified protein using hydrazide-agarose. Subsequent incubation of the immobilized protein using a fluorescently labeled or PEGylated alkoxyamine resulted in the release of pure GFP containing the desired site-specific covalent modifications. This procedure was also employed to produce PEGylated glucose-dependent insulinotropic polypeptide[137] (GIP), a protein with potential therapeutic activity for diabetes.[138]

Experimental Section

Enzymatic studies of FPP-analogues 1 and 2 using a continuous fluorescence assay. Enzymatic reaction mixtures contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (20 mM), $ZnCl_2$ (10 μM), 2.4 μM *N*-dansyl-GCVIA (**3**), 0.04 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 80 nM PFTase, and varying concentrations of either **1** or **2** (0-50 μM), in a final volume of 250 μL. The reaction mixtures were equilibrated at 30 °C for 5 min, initiated by the addition of PFTase, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 10 min. The initial rates of formation of products were obtained as slopes in IU/min using least squares analysis. Corrections were applied to all of the rate calculations based on the difference between the fluorescence intensity of the prenylated product and the starting peptide. Assuming 100% conversion, the difference corresponds only to the fluorescence of the total amount of the product. The slope was then divided by the fluorescence difference followed by multiplying by the total concentration of peptide (2.4 μM) which then gives the rate of formation of product

in μM/s. It should be noted that the K_M values reported here are actually apparent K_M values, since the measurements were performed at only a single peptide concentration. The data were fit to a Michaelis-Menten model using a nonlinear regression program, to determine k_{cat} and K_M.

Enzymatic synthesis of 4a and 5a. Enzymatic reactions (26 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (20 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), **3** (2.4 μM), PFTase (80 nM), and either **1** or **2** (30-50 μM). To ensure complete disulfide reduction of the peptide, all reagents except substrates and enzyme were pre-mixed and incubated, for 2 h at 4 °C. With all reagents mixed, the reaction was initiated by the addition of enzyme and the resulting mixture was incubated at 30°C for 1 h. The reaction progress was monitored by UV absorbance (λ=340 nm, absorbance of the dansyl chromophore) using analytical RP-HPLC. The following conditions were employed: flow rate: 1 mL·min^{-1}; 500 μL injection loop; gradient 0–100% B in 30 min; solvent A: NH_4HCO_3 (25 mM in H_2O); solvent B: CH_3CN. After 1 h, the reaction was purified by using a Waters Sep-Pak Plus reversed-phase C_{18} Environmental Cartridge. The cartridge was first washed with solvent B (10 mL) then equilibrated with solvent A (20 mL). The crude enzymatic reaction mixture was applied to the cartridge and a gradient elution was performed in the following sequence: 10 mL solvent A, 10 mL solvent C (20% solvent B, 80% solvent A), 10 mL solvent D (40% solvent B, 60% solvent A), 10 mL solvent E (60% solvent B, 40% solvent A). Fractions (1 mL per tube) were collected and the product elution was monitored using a handheld UV lamp. The green-fluorescent product was clearly visible and the brightest fraction was selected and its purity was confirmed by

RP-HPLC. LC-MS analysis of the purified products gave ions of 913.5 and 979.4 as the predominant species, which are consistent with $[M+H]^+$ for **4a** and **5a,** respectively.

Oxime ligation between peptide-aldehyde 4a and 5a and aminooxy alexafluor-488 (6c). Coupling reactions contained 3-5 µM **4a** or **5a**, 200 µM alexafluor-488 (**6c**), PB (0.1 M, pH 7.0), and aniline (100 mM) in a final volume of 500 µL. Reactions were performed at rt and were initiated by addition of aniline (100 mM). LC-MS analysis of the reaction mixture after 3-4 h gave ions of 1384.6 and 1450.6 as the predominant species, which are consistent with $[M+H]^+$ for **4c** and **5c,** respectively.

Hydrazone ligation between peptide-aldehydes 4a and 5a and Texas red hydrazide (6b). Coupling reactions contained 3-5 µM **4a** or **5a**, 200 µM Texas red (**6b**), PB (0.1 M, pH 7.0), and aniline (100 mM) in final volume of 500 µL. Reactions were performed at rt and were initiated by addition of aniline (100 mM). LC-MS analysis of the reaction mixture after 1 h gave ions of 758.46 and 791.45 as the predominant species, consistent with $[M+2H]^{2+}$ for hydrazones **4b** and **5b**, respectively.

Enzymatic incorporation of compounds 1 and 2 into GFP-CVIA (7).[80] Enzymatic reaction mixtures (10 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 µM), DTT (5.0 mM), **7** (2.4 µM), either **1** or **2** (30-50 µM), and PFTase (80-200 nM). After incubation at 30°C for 2 h for **1**, and overnight for **2**, the respective reaction mixtures were concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess of **1** or **2** was removed through a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluant. The

subsequent protein concentration was calculated by UV absorbance at 488 nm (ε=55,000 $M^{-1}\cdot cm^{-1}$).

Coupling reaction between aldehyde-labeled GFP-CVIA (8a and 9a) with alexafluor-488 (6c). Alexafluor-488 (**6c**) (3.2 µL of 3.2 mM solution in DMSO) was added to 42 µL of **8a** or **9a** (stock solution of 60 µM in PB). PB (2 M, pH 7.0, 2.5 µL) was added and the reaction was initiated by adding aniline (100 mM) and was allowed to proceed for 3-5 h at rt. The mixture was then purified by a NAP-5 column to remove excess dye. LC-MS analysis of the sample showed only oxime ligated protein and no free aldehyde was detected indicating a complete reaction in both cases.

Immobilization of 9a onto hydrazide agarose beads. Hydrazide agarose beads (Thermo Scientific, hydrazide loading: 16 µmol/mL) (300 µL) were washed with PB (0.1 M, pH 7.0, 3x500 µL). PB (30 µL, 1 M, pH 7.0) was added to the beads followed by addition of **9a** (200 µL, 87 µM). Immobilization was initiated by adding aniline (2 µL, 100 mM). For controls, GFP-CVIA (**7**) was added instead of **9a**. The solution was centrifuged, then the GFP UV-absorbance of the supernatant was measured as a function of time (488 nm, ε=55,000 $M^{-1}\cdot cm^{-1}$). After 2 h, the solution was centrifuged and the beads were washed thoroughly with PB (0.3 M, pH 7.3, 3x300 µL) and KCl (1 M, 3x300 µL) to remove non-specifically bound proteins and were stored in pH 7.5 Tris buffer at 4 °C.

Release of immobilized GFP from beads using hydroxylamine. The GFP-beads were incubated in PB (0.3 M, pH 7.0) with hydroxylamine (200 mM) and aniline (100 mM) and the resulting mixture was vortexed. The solution was centrifuged and then the UV-

absorbance of the GFP in the supernatant was measured as a function of time (488 nm, ε=55,000 $M^{-1} \cdot cm^{-1}$).

Coupling reaction between aldehyde-labeled GFP-CVIA (8a and 9a) with alexafluor-488 (6c). Alexafluor-488 (**6c**) (3.2 µL of 3.2 mM solution in DMSO) was added to 42 µL of **8a** or **9a** (stock solution of 60 µM in PB). PB (2 M, pH 6.7, 2.5 µL) was added and the reaction was initiated by adding 100 mM aniline and was allowed to proceed for 5-6 h at rt. The mixture was then purified using a NAP-5 column to remove excess dye. LS-MS analysis of the sample showed only oxime-ligated protein and no free aldehyde was detected indicating a complete reaction in both cases.

Coupling reaction between aldehyde-labeled GFP-CVIA (8a and 9a) with Texas red hydrazide (6b). Texas red hydrazide (**6b**) (7 µL of 1.9 mM solution in DMSO) was added to 100 µL **8a** and **9a** (stock solution of 40 µM in PB). The reaction was initiated by adding 100 mM aniline and was allowed to proceed for 1 h at rt. The mixture was then purified using a NAP-5 column to remove excess of **6b**. LC-MS analysis of the sample showed the presence of both hydrazone ligated proteins **8b** and **9b**, and the free aldehydes **8a** and **9a**. The ratios of free aldehydes to their respective hydrazone products were ~4, indicating only ~20% completion within this range of reactant concentrations. The mass spectrum in Figure 3.2E which corresponds to conversion of compound **8a** to **8b** was made by superimposing the two separate mass spectra of both the free aldehyde protein **8a** and the hydrazone ligated protein **8b** present in the product mixture, based on their relative intensities. The aldehyde- and hydrazone-functionalized proteins have

different retention times and thus show two different peaks in the corresponding LC chromatograms.

FRET studies between GFP-aldehyde 9a and Texas red hydrazide (6b). Texas Red hydrazide (**6b**) (7 µL of 1.9 mM solution in DMSO) was added to 90 µL of **9a** (stock solution of 60 µM in PB, 0.1 M, pH 6.7). The reaction was initiated by adding 0.9 µL aniline (100 mM) followed by vortexing the solution and allowing it to proceed for 1 h at rt. The mixture was then purified using a NAP-5 column to remove excess of **6b**. The protein solution was diluted to 1 nM and its fluorescence was measured. For the first control, the same amount of GFP-CVIA (**7**) fluorescence was measured and compared with that of **9b**. As a second control, the ligated protein (**9b**) was heated for few minutes to denature the protein and the resulting fluorescence was measured to verify that FRET required both the protein and the Texas red fluorophores.

FRET studies between GFP-aldehyde 9a and aminooxy-TAMRA (6d). Aminooxy-TAMRA (**6d**) (100 µM) was added to 90 µL of **9a** (stock solution of 50 µM in Tris·HCl (100 mM, pH 7.0). The reaction was initiated by adding 0.9 µL aniline (100 mM) followed by vortexing the solution and allowing it to proceed for 3 h at rt. The mixture was then purified using a NAP-10 column to remove excess of **6d**. The collected solution appeared red and not green suggesting that efficient FRET was occurring between the GFP and TAMRA fluorophores. The fluorescence of the solution was measured and compared with two other controls. For the first control, the same amount of GFP-CVIA (**7**) fluorescence was measured and compared with that of **9e**. As a second control, the

same amount of TAMRA fluorescence was measured. The two controls confirmed that FRET was occurring between the protein and the TAMRA fluorophores.

Crude prenylation, immobilization and subsequent labeling and release of GFP-CVIA. A pellet of cells expressing GFP-CVIA were suspended in buffer (20 mM Tris·HCl pH 7.5, 1 mM EDTA), sonicated and clarified by centrifugation. The GFP concentration present in the crude soluble protein mixture was calculated by UV absorbance at 488 nm. Next, prenylation was performed by adding PFTase (200 nM), **2** (50 µM), Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 µM) and DTT (5.0 mM) to a solution of **7,** to achieve a final concentration of 2.0 µM in the crude mixture. After overnight incubation at 30°C, the reaction mixture was filtered and concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess **2** was removed through a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluting solvent. The subsequent GFP concentration in the crude mixture was calculated by UV absorbance at 488 nm and was determined to be 30 µM. Immobilization was performed as described above. The beads were washed thoroughly with PB (0.3 M, pH 7.3) and KCl (3x300 µL, 1 M) to remove non-specifically bound proteins followed by incubation with aminooxy fluorophore **6c** (1 mM) and aniline (100 mM) overnight with constant agitation. The supernatant was then analyzed via SDS-PAGE and in-gel fluorescence analysis to confirm the labeling and release of the protein from the beads.

Coupling reaction between aldehyde-labeled GFP-CVIA (9a) with aminooxy PEG (10). Aminooxy PEG (**10**) (1.5 mg, MW 10 kDa) was added to 100 µL **9a** (stock solution

of 10 μM in 50 mM Tris·HCl). PB (pH 7) was added to a final concentration of 0.1 M. The reaction was initiated by adding 100 mM aniline and was allowed to proceed for 1 h at rt. SDS-PAGE analysis of the sample was used to confirm covalent attachment of the PEG **10** to aldehyde **9a**. Excess **10** and PB were removed using a zip-tip protocol followed by MALDI MS analysis of the sample to characterize the product and demonstrate that no free aldehyde was present indicating complete reaction.

PEGylation from immobilized GFP-beads. Immobilization was performed as described above. Beads were washed thoroughly with PB (0.3 M, pH 7.3, 3x300 μL) and KCl (1 M, 3x300 μL) to remove non-specifically bound proteins. Next, the beads were incubated with aminooxy PEG **10** (2 mM) and aniline (100 mM) overnight while vortexing the solution. SDS-PAGE analysis of the supernatant indicated the successful PEGylation and release of the aldehyde-GFP from the hydrazide-beads.

Enzymatic prenylation of GIP-CVIM (12a) with aldehyde substrate 2. Enzymatic reaction mixtures (10 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), **12a** (2 μM), **2** (50 μM), and PFTase (200 nM). After incubation at 30°C overnight, significant precipitate was present in the solution, suggesting possible precipitation of GIP was occurring upon prenylation. The precipitate was separated from the solution by centrifugation at 12,000xg for 15 min, washed with 25 mM $(NH_4)_2CO_3$ buffer to remove excess **2,** and centrifuged again. MALDI-MS analysis of the precipitate (dissolved in H_2O, 0.5% TFA, v/v) confirmed the prenylation of GIP with **2,** while the solution showed neither starting GIP **12a** nor the prenylated material **12b**.

Coupling reaction between aldehyde-labeled GIP (12b) with aminooxy PEG (13). A small amount of the precipitate **12b** was dissolved in H_2O containing 0.5% TFA (v/v) and aminooxy PEG (**13**) was added to a final concentration of 200 µM. The reaction was allowed to proceed for 2 h at rt. Excess **13** was removed using a zip-tip protocol. MALDI-MS analysis of the sample was employed to characterize the product and to demonstrate that no free aldehyde was present indicating complete reaction.

Prenylation of GIP (12a) in crude E. coli extract. *E. coli* extract containing GIP-CVIM (**12a**) was subjected to enzymatic prenylation by incubating it in the presence of PFTase (200 nM), **2** (50 µM), Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 µM) and DTT (5.0 mM). After overnight incubation, the precipitate was separated from the solution by centrifugation at 12,000xg for 15 min, washed with 25 mM $(NH_4)_2CO_3$ buffer to remove excess of **2,** and then centrifuged again. MALDI-MS analysis of the precipitate, performed as described above, was used to confirm the prenylation of GIP with **2**.

Immobilization and subsequent PEGylation and release of resin-bound GIP. Immobilization was performed as described above for GFP except that GIP was dissolved in H_2O containing 0.5% TFA (v/v). No aniline catalyst was added to the solution in this case. After incubation for 1 h, the beads were washed thoroughly with H_2O (3x300 µL) to remove non-specifically bound proteins from the beads. Next, the beads were incubated with aminooxy PEG **13** (~400 µM) in H_2O containing 0.5% TFA overnight with constant agitation of the solution. Excess **13** was removed using a zip-tip protocol. MALDI MS

analysis of the sample was employed to confirm the presence of the desired product and to assess the purity of the PEGylated GIP (**14**).

General procedure for MALDI analysis of protein samples. A zip-tip (C_4 column) was first washed with 10 µL solvent A (CH_3CN containing 0.1% TFA; v/v) and then equilibrated with solvent B (H_2O containing 0.1% TFA; v/v). The sample (10 µL) was then adsorbed onto the C_4 matrix via repeated cycles of aspiration and ejection (5-10 cycles) using a pipettor. Next, the zip-tip was washed 5x10 µL with solvent B and the proteins eluted with 2 µL of a mixture of solvent A and B (75:25). Next 0.7 µL of the eluted material was added to a MALDI plate and 0.7 µL of matrix was added on top of the sample plate to form crystals. A saturated solution of sinapinic acid (3,5-dimeth-oxy-4-hydroxy-cinnamic acid) was used as the matrix.

Results and Discussion.

Comparison of alkyl and aryl aldehydes as substrates for PFTase. To examine the ability of PFTase to be used in a protein modification strategy employing oxime and hydrazone formation, we first wanted to explore the range of aldehydes that could be accepted as alternative substrates for PFTase. Thus, compound **2**, containing an aryl aldehyde was designed and synthesized in five steps from geraniol (Scheme 3.S2). In brief, THP-protected geraniol was initially oxidized at C-8 to a terminal alcohol,[139] followed by acylation with formylbenzoic acid using EDC as the coupling reagent. The THP group was removed and the alcohol was converted to the corresponding allylic bromide using CBr_4 and PPh_3. Subsequent displacement with $[(n\text{-}Bu)_4N]_3HP_2O_7$ followed by purification via ion-exchange chromatography and RP-HPLC yielded product in

which the desired aldehyde was almost completely transformed to the corresponding carboxylic acid; therefore, a direct phosphorylation strategy using $(HNEt_3)_2HPO_4$ and CCl_3CN as the activating reagent was employed. Subsequent purification by RP-HPLC produced the desired aldehyde analogue **2** in 5.4% overall yield whose structure was confirmed by ^{1}H-NMR, ^{31}P-NMR,[140] and HR-ESI-MS. Aldehyde **1** was prepared in six steps starting from farnesol as previously described[82] with several modifications that significantly improved the overall yield to 1.3%. Despite these improvements, the synthesis of **2** was significantly more efficient primarily due to the selectivity in the SeO_2 oxidation step. The preparation of **1** proceeds via THP-protected farnesol whereas the synthesis of **2** uses THP-geraniol. Selective oxidation of the alkene at C-6 (over the electron poor C-2 alkene) in geraniol is facile compared to the preferential oxidation of the alkene at C-10 in farnesol due to competing reaction with the C-6 olefin which exhibits comparable reactivity. Hence, the reaction cannot be driven to completion resulting a significantly reduced yield (compare 56% for geraniol oxidation to 23% for farnesol oxidation).

Initially, prenylation reactions containing *N*-dansyl-GCVIA (**3**), **2**, and PFTase were monitored by HPLC and LC-MS/MS. As was observed previously with **1**, a new species with longer retention time appeared in the reaction mixture containing **2**. LC-MS analysis of that compound gave an $[M+H]^+$ peak at 979.4 Da, consistent with the proposed structure of peptide **5a** (Supporting Information). Next, a kinetic analysis of the incorporation of analogue **2** by PFTase was performed using a continuous fluorescence-based enzyme assay as had previously been carried out with **1**. Varying concentrations of

2 were incubated with the fluorescent peptide substrate, *N*-dansyl-GCVIA, and PFTase; the rates of those enzymatic reactions were determined and shown to obey saturation kinetics. Steady-state kinetic parameters for prenylation reactions with the two aldehyde analogues are summarized in Table 3**.**1 with additional details provided in the Supporting Information section (Figure 3.S2). Comparison of the catalytic efficiencies for these alternative substrates indicates that both compounds have reduced efficiency relative to FPP, manifesting k_{cat}/K_M values of 0.23 and 0.05, respectively (relative to FPP). We found that decreases in k_{cat} constituted the major reason for the diminished catalytic efficiency of the analogues; k_{cat} for aldehyde **1** was 4-fold lower while k_{cat} for aldehyde **2** was 35-fold lower (compared to k_{cat} for FPP). No significant differences were observed in the K_M values for the different analogues. Thus, in summary, while **1** is the superior alternative substrate, **2** is easier to prepare making these two compounds functionally interchangeable.

Preparation and reactivity of PFTase-mediated aldehyde-functionalized peptides. In our earlier work with **1**, experiments with the aldehyde functionalized peptide **4a** focused on oxime-forming reactions. Here, we sought to expand the scope of possible chemistry to include hydrazone formation as well. Accordingly, large scale (26 mL) reactions containing *N*-dansyl-GCVIA (**3**), **1** or **2**, and PFTase were performed and the products isolated after purification via solid phase extraction. The resulting material was subsequently used to evaluate ligation reactions between Texas red hydrazide **6b**/alexafluor-488 aminooxy **6c** and aldehyde-containing peptides **4a** and **5a** in the presence of aniline**.**[38] Kinetic analysis of oxime formation showed that in the range of 2-4

μM of **4a** and **5a**, ligations at pH 7 were essentially complete within 3-4 h. LC-MS analysis of the reaction mixture resulted in $[M+H]^+$ peaks being observed at 1384.6 and 1450.6 Da, consistent with the production of oximes **4c** and **5c**, respectively (Supporting Information). In contrast, hydrazone formations with **4a** and **5a** in the same concentration range of reagents showed only 30-50% completion but within 30-60 min (a significantly shorter time frame). LC-MS analysis of the reaction mixtures resulted in $[M+2H]^{2+}$ peaks observed at 758.46 and 791.45 Da, consistent with the formation of hydrazones **4b** and **5b**, respectively (Supporting Information). Overall, these experiments with aldehyde-containing peptides **4a** and **5a** suggest that hydrazone ligations have the advantage over oxime-forming reactions of reaching equilibrium at a higher rate but at the cost of lower conversion to the conjugated products due to their lower association constants.

Preparation and reactivity of PFTase-mediated aldehyde-functionalized proteins. With the ability of aldehyde analogues **1** and **2** to be incorporated by PFTase and their subsequent derivatization via oxime and hydrazone ligations established in a peptide model system, we next evaluated the utility of the aldehyde analogues for selective protein modification. Accordingly, aldehydes **1** and **2** were incubated with GFP-CVIA

Table 1. Steady-state kinetic parameters of substrates, and HPLC Retention times for prenylated peptide products.

Compound[a]	k_{cat} (s^{-1})	K_M (μM)	$(k_{cat}/K_M)_{rel}$[a]	Rt (min)
FPP	0.52	1.71	1	--
1	0.133±0.003	1.87±0.17	0.23	--
2	0.015±0.001	1.02±0.16	0.05	--
N-dansyl-	--	--	--	21.5
GC(Far)VIA	--	--	--	18.9
4a	--	--	--	18.6
5a				

[a] V_{rel} refers to k_{cat}/K_M with respect to FPP

(**7**) in the presence of PFTase for 2 h, and overnight, at 30 °C, respectively. Those reaction times were based on our earlier observations that peptide substrate **3** could be prenylated withaldehyde analogues **1** and **2** in less than 1 h, and 4 h, respectively.

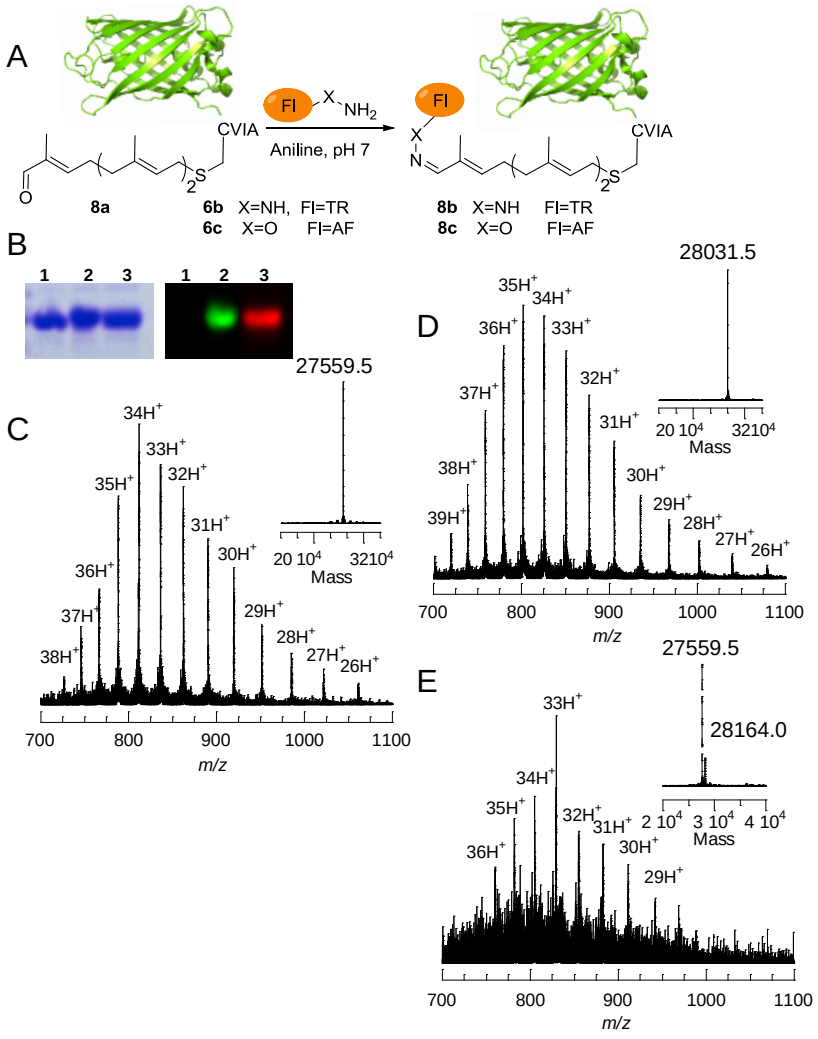


Figure 3.2. A) Schematic representation of oxime and hydrazone ligations of **8a** to yield **8b** and **8c**. B) Fluorescence (right) and Coomassie blue staining (left) images of a gel loaded with **8a** labeled with alexafluor **6c** and Texas red **6b** via oxime and hydrazone ligations, respectively, showing covalent attachment of fluorophores to the protein. lane 1: GFP-CVIA **7**; lane 2: **8c**; lane 3: **8b** C, D and E) ESI MS spectra of **8a** (spectrum C) and hydrazone/oxime ligation products **8b** and **8c**, showing full conversion for oxime

(spectrum D) and ~20% for hydrazone (spectrum E) ligations with the deconvoluted mass spectra shown in the insets.

Concentration by ultracentrifugation followed by size-exclusion chromatography to remove unreacted substrates yielded aldehyde-functionalized GFP-CVIA **8a** and **9a**. Reaction completion was confirmed by LC-MS analysis (Figure 3.1C and 3.1D), in which none (in the case of **8a**) or very small amounts (in the case of **9a**) of free GFP-CVIA (**7**) could be detected in comparison to the large peaks for prenylated GFPs. Deconvolution of the LC-MS data from the purified protein products showed species at 27,559.0 Da and 27,625.5 Da, consistent with the structures of aldehyde-GFPs **8a** and **9a**. In general, LC-MS analysis of GFP and its congeners has proved to be quite powerful for studying these reactions. As noted above, in a preliminary communication,[82] we had shown that aldehyde-GFPs **8a** could be derivatized to produce oxime-linked products. Here, it was desired to expand those experiments to include hydrazone formation and to compare the relative reactivity between the two different aldehyde donors, **8a** and **9a**, containing α,β-unsaturated- and aryl-aldehydes, respectively. To fluorescently label those aldehyde-functionalized proteins, we chose Texas red hydrazide (**6b**) and Alexaflour-488 aminooxy (**6c**) for their excellent quantum yields and high visible light absorbtion. Thus, aldehyde-GFPs **8a** and **9a** were incubated separately with alkoxyamine **6c** at pH 7 and rt. Kinetic analysis, performed via LC-MS measurements, showed that the reaction required 3-4 h to proceed to completion. At that point, no detectable unmodified protein-aldehydes (**8a** and **9a)** were observed. Gratifyingly, the deconvoluted MS data indicated the presence of species at 28,032.0 Da and 28,120.5 Da, consistent with the proposed oximes

8c and **9c**. In-gel fluorescence analysis performed under denaturing conditions confirmed covalent attachment of aminooxy **6c** to the aldehyde-containing proteins (Figure 3.2B). Unprenylated GFP-CVIA (**7**) failed to show any labeling with alkoxyamine **6c**, further confirming that the ligations require the presence of the enzymatically introduced aldehyde functionality and that the ligation reaction is truly bioorthogonal. Overall, the oxime ligation reactions appear to be highly efficient since no unligated aldehyde-GFPs (**8a** or **9a**) were observed upon LC-MS analysis (Figure 3.2C and 3.2D) of the ligation reaction mixtures.

Aldehyde-functionalized GFPs **8a** and **9a** were also each incubated with hydrazide **6b** at pH 7 and rt under the same conditions employed in the aforementioned oxime ligations. LC-MS analysis of aldehyde-functionalized GFP-containing reactions after 1 h showed approximately 20% conversion of aldehydes **8a** and **9a** to their respective hydrazones **8b** and **9b** (Figure 3.2C and 3.2E); more extensive reaction times did not result in the appearance of additional hydrazone product suggesting that the reaction had reached equilibrium within 1 h. These results are in good agreement with those from hydrazone ligations for aldehyde-functionalized peptides **4a** and **5a** described above.

Application to FRET analysis of labeled GFP. To demonstrate the utility of this method for applications beyond simple protein labeling, we next investigated the ability of Texas red-labeled GFP (**9b)** to undergo fluorescence resonance energy transfer (FRET). After performing the ligation reaction between aldehyde **9a** with Texas red hydrazide **6b** at rt for 1 h, excess fluorophore was removed via size exclusion

chromatography. A strong fluorescent signal at 640 nm (emission wavelength of Texas red) was observed upon excitation at 488 nm (excitation wavelength of GFP), indicative of FRET between Texas red and GFP due to their close proximity resulting from covalent attachment of the fluorophore to the protein (Figure 3.3). When the hydrazone-ligated protein was denatured, no FRET was observed upon excitation at 488 nm (Figure 3.3, red spectrum) and only a small background peak was observed upon excitation of GFP that had not been modified with Texas red (Figure 3.3, green spectrum), further confirming that FRET was occurring between the aldehyde- functionalized protein and the fluorophore. While the above results appeared promising, the FRET efficiency could not be calculated from the experimental data due to incomplete hydrazone ligation

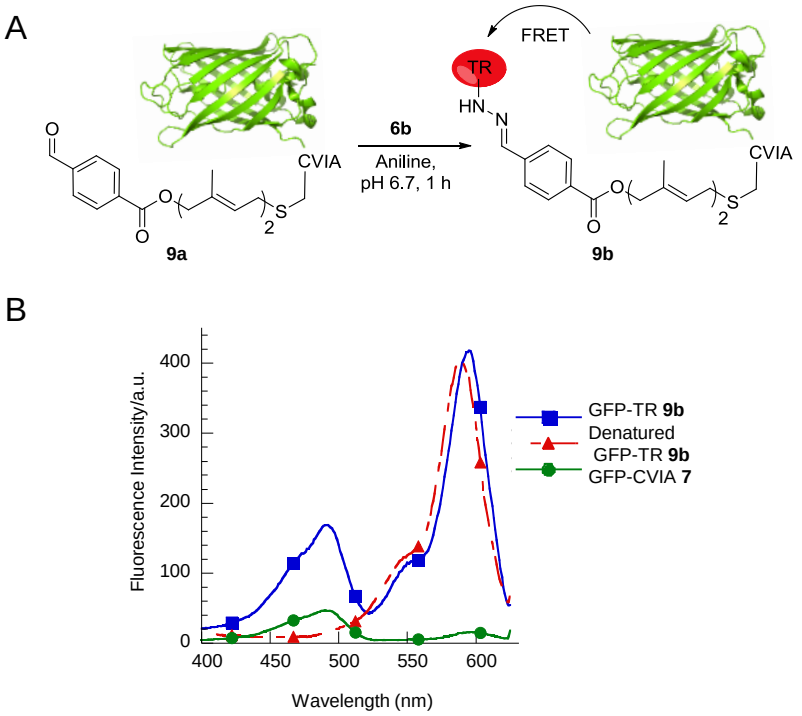


Figure 3.3. A) Schematic representation of the fluorescent labeling of **9a** via hydrazone ligation. The conjugated protein was expected to show FRET between Texas red and GFP-CVIA. B) Excitation spectra obtained by monitoring at 640 nm. Squares: **9b**; Triangles: denatured **9b**; Circles: **7**; all three samples had equal concentrations of the chromophores.

reaction. Hence, aminooxy-TAMRA **6d** was ligated with aldehyde-protein **9a**. As expected, LC-MS analysis of the oxime formation reaction mixture showed a peak at 28,111 Da consistent with the structure of TAMRA labeled GFP **9d** and showed no **9a**, in good agreement with the high efficiency observed in previous oxime ligation reactions. Emission spectra of **9e**, monitored at 488 nm excitation, showed FRET while the same amount of GFP-CVIA (**7**) and fluorophore **6d** showed a substantially larger emission band at 510 nm and a smaller band at 580 nm, respectively. Energy was transferred from

donor (GFP) to acceptor (TAMRA) with an efficiency greater than >96% and the distance was calculated to be 37 Å (Figure 3.6B), consistent with a distance of 35 Å

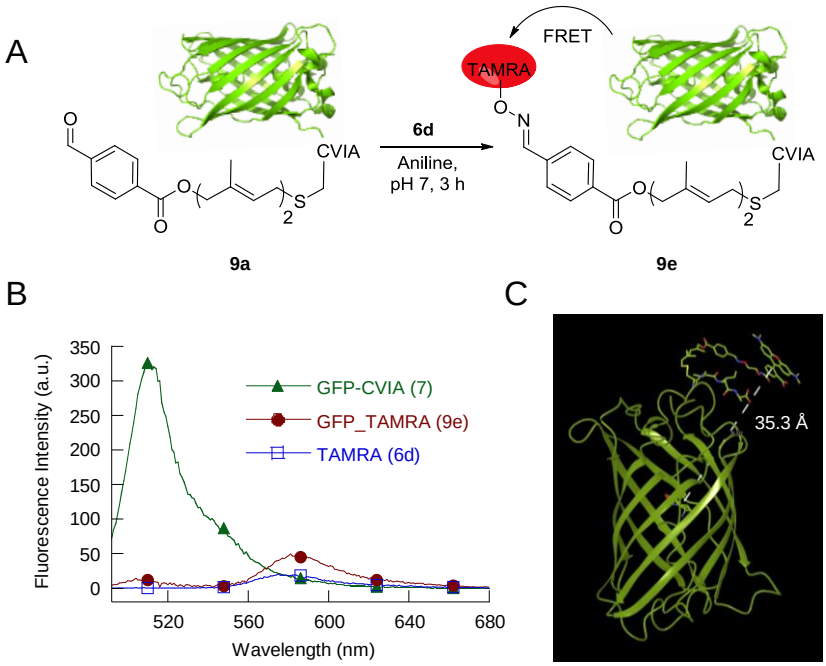


Figure 3.4. A) Schematic representation of the fluorescence labeling of **9a** via oxime ligation. The conjugated protein was expected to show FRET between TAMRA and GFP. B) Emission spectra obtained by excitation at 488 nm. Circles: **9e**; Triangles: GFP-CVIA (**7**); Squares: **6d**; all three samples had equal concentrations of the chromophores. C) Molecular model of GFP–TAMRA **9e** conjugate.

calculated for the model GFP–TAMRA (Figure 3.6C) and measured from the GFP chromophore to the TAMRA fluorophore (See the Supporting Information for a description of the modeling).

Reversible immobilization of purified aldehyde-functionalized GFP using hydrazide-modified agarose beads. Next, we examined two additional applications for the aldehyde-functionalized proteins described here. First, their utility in protein

immobilization was examined (Figure 3.5). Hydrazide- functionalized agarose beads were incubated with aldehyde-GFP **9a** at rt in the presence of 100 mM aniline. The immobilization reaction was followed by monitoring the UV absorbance at 488 nm of the supernatant as a function of time. Results from those measurements showed that equilibrium was reached in approximately 45 min; in that time, the beads became highly fluorescent (Figure 3.6A); less fluorescent

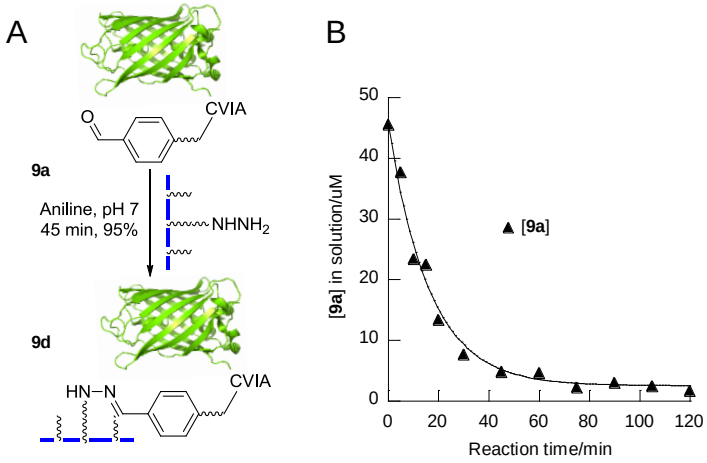


Figure 3.5. A) Schematic representation of immobilization of **9a** onto hydrazide functionalized agarose beads to yield **9d**. B) Kinetic analysis of immobilization of **9a** onto hydrazide functionalized agarose beads. The reaction was carried out at rt, in the presence of 100 mM aniline and excess beads. UV absorbance of GFP in the supernatant was measured at different times showing >95% immobilization in ~45 min. The data was fit to a simple exponential process.

beads were observed in the absence of aniline catalyst and no fluorescent beads were seen using GFP lacking the aldehyde moiety (Figure 3.6C). Based on the amount of aldehyde-GFP **9a** remaining in the supernatant, the efficiency of covalent immobilization was calculated to be greater than 95% (Figure 3.5B), an impressive result for site-specific protein immobilization. Next, oxime ligation using hydroxylamine in the presence of

aniline was employed to remove the covalently immobilized hydrazone-GFP **9d**. Hydroxylamine (200 mM) was incubated with **9d** in presence of aniline (100 mM) at rt and the UV absorbance at 488 nm of the supernatant was measured as a function of time. In this case, analysis of the results showed that in approximately 3 h, 80% of the immobilized GFP was released from the beads, and accordingly, the beads became significantly less fluorescent (Figures 3.6B and 3.7). For comparison, the hydrolytic

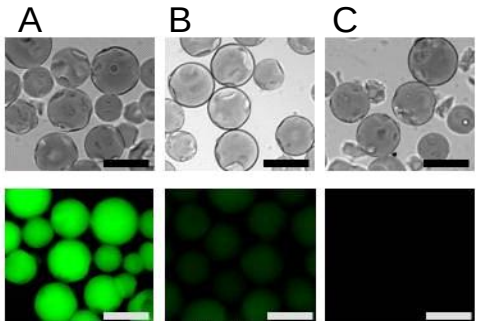


Figure 3.6. Immobilization onto and subsequent release of **9a** from hydrazide-functionalized agarose beads: A) immobilization reaction mixture in the presence of aniline, B) release of **9d** from agarose beads via oxime ligation with hydroxylamine in the presence of aniline for ~3 h, and C) control immobilization reaction containing unmodified GFP-CVIA **7**. The immobilization reaction was carried out in the presence of protein (54 μM), aniline (100 mM) and PB (100 mM, pH 7). Release of hydrazone-GFP **9d** from agarose beads was carried out in the presence of hydroxylamine (200 mM), aniline (100 mM) and PB (200 mM, pH 7). Bright-field images are on the top and fluorescent microscope images are on the bottom. Scale bars in the lower right-hand corners represent 200 μm.

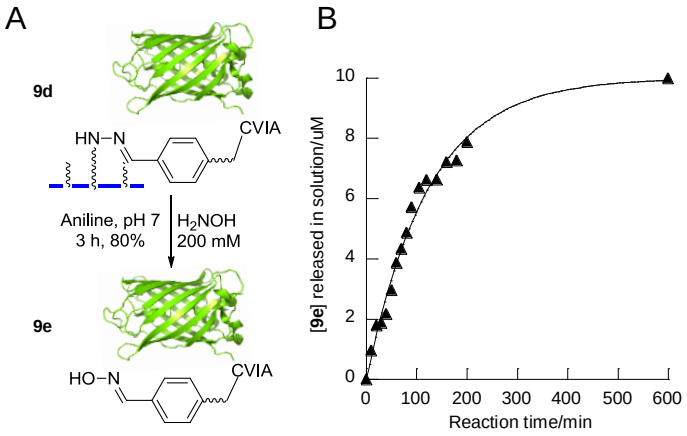


Figure 3.7. A) Schematic representation of the release of immobilized GFP **9d** to yield **9e** from agarose beads via oxime ligation with hydroxylamine. B) Kinetic analysis of the release of **9e** from agarose beads by oxime ligation. The reaction was carried out at rt, in the presence of 100 mM aniline and 200 mM of hydroxylamine. UV absorbance of GFP in the supernatant was measured over time, which showed approximately 80% release of **9e** in 3 h. Analysis of the hydrolytic stability of **9d** in the absence of hydroxylamine and aniline showed no detectable release of GFP on the same time scale. The data was fit to a simple exponential decay process.

stability of immobilized GFP in the absence of hydroxylamine and aniline was also analyzed and the results showed that the hydrazone bond in pH 7.5 Tris buffer was completely stable for 48 h with no detectable release of GFP.[141] This achievement highlights a significant advantage of this chemistry over click chemistry and other irreversible methods since it can be used to efficiently covalently immobilize proteins onto solid surfaces and then release them under mild conditions without protein denaturation. Addition of aniline catalyzes the hydrolysis of hydrazone to hydrazide and aldehyde. Since oxime formation has a larger equilibrium constant than hydrazone formation,[142] the presence of hydroxylamine and aniline drives the equilibrium from hydrazone towards oxime formation and free hydrazide.

Enzymatic modification, immobilization and labeling in crude extract. An important feature of the labeling method described here is that it uses an enzymatic process for the introduction of aldehyde groups into proteins. Due to the specificity of that biocatalytic process and the fact that there are no endogenous proteins in *E. coli* that contain a C-terminal CAAX box sequence, we reasoned that it should be possible to selectively functionalize proteins present in crude extract without purification. Additionally, once modified, it should also be possible to immobilize aldehyde-containing proteins and release them with an alkoxyamine that includes a fluorophore or PEG chain. In that way, a single protein present in *E. coli* crude extract could be modified, immobilized and labeled without purification. To explore this, *E. coli* cells expressing GFP-CVIA were grown, lysed, and subjected to enzymatic prenylation using PFTase and substrate **2**. LC-ESI/MS analysis of the reaction mixture was employed to confirm the introduction of the aldehyde functionality into GFP-CVIA **7** in the crude cell lysate. The reaction mixture was then concentrated and excess of **2** was removed via size exclusion column chromatography (NAP-5 column). Aldehyde-GFP **9a** was then selectively immobilized from the crude cell lysate onto hydrazide-functionalized beads using aniline as the catalyst. Immobilization was followed by measuring the GFP absorbance present in the solution and was judged to be complete within 45 min at which time the beads became highly fluorescent and the supernatant solution became almost colorless. Next, the beads were washed to remove any non-specifically bound proteins and were then treated with aminooxy fluorophore

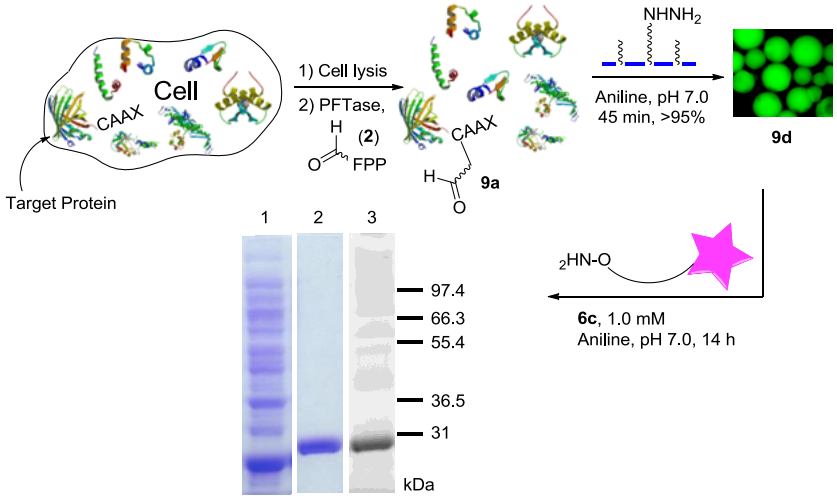


Figure 3.8. Chemoenzymatic site-specific tagging of proteins by aldehyde-FPP analogs by PFTase followed by capture of the aldehyde-functionalized protein in the crude cell lysate via hydrazide functionalized beads. Prenylation in the crude extract was confirmed by LC-MS analysis. The immobilized protein was then released into the solution or fluorescently labeled by addition of hydroxylamine or an aminooxy-fluorophore, using aniline as the catalyst. SDS-PAGE analysis: lane 1: crude *E. coli* lysate containing **9a** visualized by Coomassie blue staining; lane 2: **9c** released from hydrazide beads after treatment with **6c** and visualized by Coomassie blue staining; lane 3: **9c** released from hydrazide beads after treatment with **6c** and visualized by in gel fluorescence analysis.

6c in presence of 100 mM aniline overnight. SDS-PAGE analysis of the supernatant solution showed a single band (Figure 3.8, lane 2) migrating with an apparent mass of 29 kDa slightly higher than that of the starting GFP (due to the addition of the aminooxy moiety) consistent with the release of GFP. In-gel fluorescence analysis (Figure 3.8, lane 3) suggested that the released protein was labeled with the fluorophore **6c**; LC-MS analysis of the released protein provided additional evidence for the formation of **9c**.

Application to protein PEGylation. The attachment of polyethylene glycol (PEG) chains to proteins is the most widely used method for improving the pharmacokinetics of polypeptide-based therapeutic agents.[143–145] Current methods for PEGylation are generally nonselective and can result in a mixture of protein-PEG positional isomers with variable biological activity.[146] Site-specific methods offer a useful alternative approach for circumventing this problem of heterogeneity. Given our success in being able to incorporate a fluorescent label into a protein via the capture and release strategy described above, we decided to evaluate the utility of this approach for the preparation of

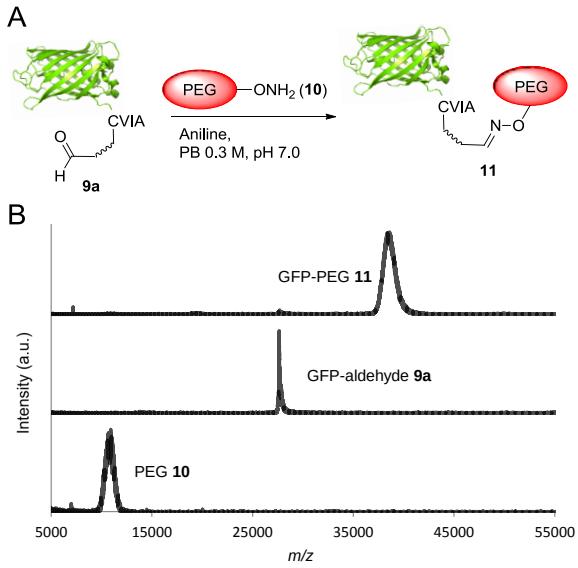


Figure 3.9. A) Generation of site-specifically C-terminal PEGylated GFP from pure **9a**. B) MALDI analysis of PEGylated GFP **11**. The lower panel is the MALDI spectrum of pure PEG **10**, the middle panel is the MALDI spectrum of pure **9a** and the top panel is the MALDI spectrum of the oxime PEGylated GFP **11**, which confirms complete conversion. The reaction was performed using **9a** (10 μM) and **10** (100 μM) for 2 h. Excess of **10** was removed via a zip-tip protocol prior to MALDI analysis.

a PEGylated protein. Thus, aldehyde- functionalized GFP **9a** was first prepared from purified GFP **7** using PFTase as described above and treated with aminooxy-PEG-10,000 **10**, to produce the protein-PEG conjugate **11**. Analysis of that material by MALDI-MS (Figure 3.9) showed an increase in molecular mass from 27.6 kDa (for **9a**) to 38 kDa for **11**; the broader peak observed for **11** is consistent with the attachment of a polydisperse polymer to a monodisperse protein. It is also important to note that no species resulting from the addition of multiple PEG chains were observed, consistent with the selective

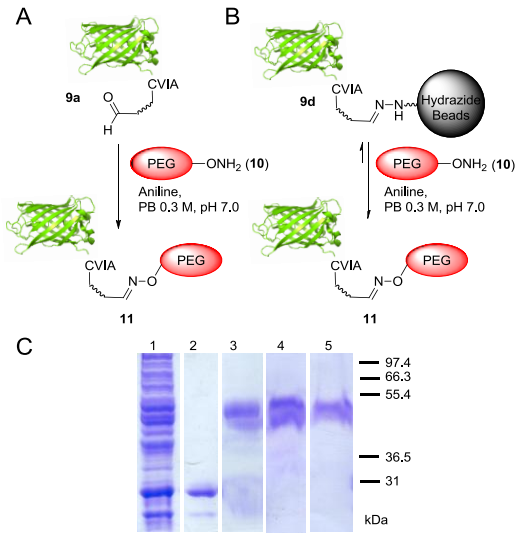


Figure 3.10. Use of PFTase-catalyzed protein modification for site-specific PEGylation from purified protein or crude cell lysate. A) Generation of site-specific C-terminal PEGylated protein from pure **9a**. B) PEGylation and release of immobilized **9d** from hydrazide beads using PEG **10**. C) SDS PAGE analysis of PEGylated GFP (**11**) from purified **9a** or from immobilized protein **9d**. In case of the crude cell lysate, **7** was chemoenzymatically and site-specifically tagged by aldehyde-containing analog **2** via PFTase catalyzed reaction, followed by capture of the resulting aldehyde-functionalized protein from the lysate using hydrazide functionalized beads. The immobilized protein was then released back into solution and simultaneously site-specifically PEGylated by addition of aminooxy-PEG **10**, using aniline as a catalyst. SDS-PAGE analysis: lane 1: crude *E. coli* lysate containing **9a**; lane 2: purified **9a**; lane 3: **11** produced by PEGylation of pure **9a** with **10**; lane 4: **11** prepared from **9d** (obtained using purified **9a**) and

subsequently released with **10**; lane 5: **11** prepared from **9d** (obtained using **9a** present in crude lysate) and subsequently released with **10**.

nature of the chemistry employed here. Analysis of the PEGylation reaction mixture by SDS PAGE revealed a decrease in electrophoretic mobility of **11** (Figure 3.10, lane 3) compared to the starting protein **9a** (Figure 3.10, lane 2). As was noted in the MALDI MS, a wider band was observed for **11** relative to **9a**, again consistent with the polydisperse nature of the protein-PEG conjugate. With the production of the PEGylated product clearly established, we next focused on generating the same material from **9a** that had not been purified chromatographically. Thus, **7** was prenylated with **2** using PFTase in crude *E. coli* extract followed by capture using hydrazide-functionalized agarose. After washing the material to remove nonspecifically bound proteins, the desired PEGylated protein (**11**) was eluted via treatment with **10** in the presence of aniline. SDS PAGE analysis showed the presence of a single band (Figure 3.10, lane 5) that comigrated with the authentic product prepared from pure **9a** (Figure 3.10, lane 3).

PEGylation of glucose-dependent insulinotropic polypeptide (GIP). The incretin, glucose-dependent insulinotropic polypeptide (GIP), is secreted from intestinal K-cells in response to nutrient ingestion and acts to augment insulin secretion in the pancreas. GIP has been proposed as a potential therapeutic agent for the treatment of type 2 diabetes based on its stimulation of insulin secretion in the presence of elevated glucose levels;[147,148] however, efforts to bring GIP forward as a drug have been hampered due to its short circulating half-life. Recently, a modified form of GIP functionalized with a C-terminal mini-PEG group has shown resistance to proteolytic degradation while preserving biological activity in an obese rat model system.[138] Accordingly, having

established the utility of our method for C-terminal site-specific modification described above with a model protein, GFP, we decided to demonstrate its utility for preparing a PEGylated form of GIP, a polypeptide with clear therapeutic potential. Thus, purified GIP-CVIM (**12a**), a form of GIP engineered to contain a C-terminal CAAX box (in this case $CVIM^{149}$) was prenylated with analog **2** under conditions established above for GFP and subsequently PEGylated using a small aminooxy-functionalized PEG containing three ethylene glycol units (**13**). This shorter PEGylation reagent was employed since it is similar in length to what has previously been shown to be effective for increasing GIP stability in serum. MALDI MS analysis (Figure 3.11) confirmed the successful prenylation and PEGylation of GIP; as was noted above with GFP, both the enzymatic prenylation and subsequent chemical PEGylation proceed with essentially complete conversion. Next, we employed the capture and release strategy developed above for GFP to prepare PEGylated GIP without prior purification. GIP- CVIM (**12a**), present in crude *E. coli*

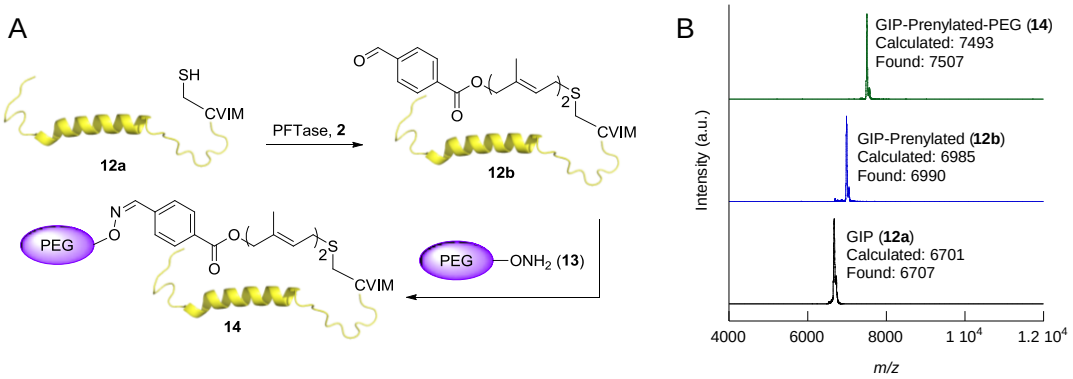


Figure 3.11. A) Schematic representation of prenylation of glucose-dependent insulinotropic polypeptide (GIP) containing a CAAX-box positioned at its C-terminus (GIP-CVIM, **12a**) with aldehyde-containing analogue **2** to yield the prenylated product **12b**, which is then site-specifically PEGylated using a short chain aminooxy-PEG (**13**). B) MALDI MS analysis of prenylation and PEGylation of GIP **12a**. MALDI MS spectra (from the top to the bottom) of oxime PEGylated GIP **14,** the prenylated aldehyde labeled GIP **12b**, and pure **12a**, respectively.

extract, was prenylated with **2** and the resulting aldehyde-functionalized polypeptide **12b** was captured on hydrazide beads. The beads were washed extensively and then treated with aminooxy-PEG **13** resulting in oxime formation and release into solution. MALDI MS analysis of the eluted material showed only the presence of PEGylated-GIP (**14**), indicating a high degree of specificity in the capture and release (Figure 3.12). Thus, this general method allows facile and effective purification of site-specifically PEGylated GIP from the crude cell extract. Overall, these experiments conclusively demonstrate how a protein, present in crude extract, can be selectively modified, labeled with a fluorophore or PEG polymer and released in pure form via a simple process that requires no

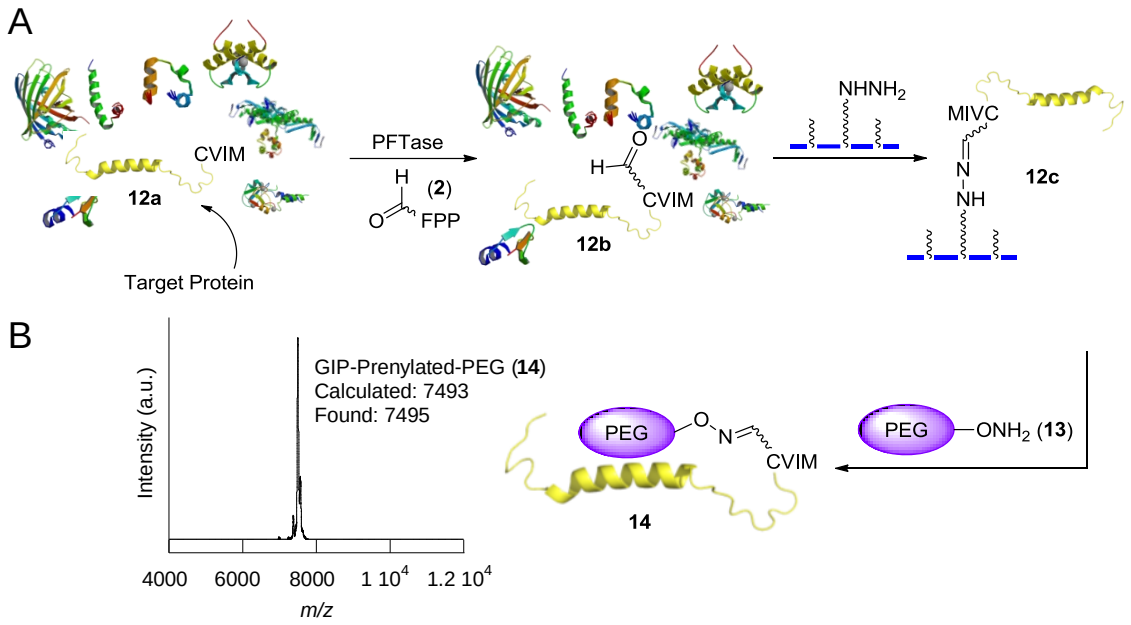


Figure 3.12. Use of PFTase-catalyzed protein modification for site-specific PEGylation of GIP **12a** from crude cell lysate. A) Chemoenzymatic site-specific tagging of GIP **12a** by aldehyde-FPP analog **2** in the crude cell lysate via PFTase followed by capture of the aldehyde-functionalized polypeptide **12b** via hydrazide functionalized beads. The immobilized polypeptide was then released back into the solution and simultaneously site-specifically PEGylated by addition of aminooxy-PEG **13**. B) MALDI analysis of the released material confirmed the formation and release of the pure PEGylated GIP (**14**) into the solution.

significant chromatographic steps. Given the specificity of the PFTase-catalyzed reaction coupled with the ability to introduce a CAAX-box onto almost any protein, this method shows great potential as a general approach for the selective immobilization and labeling of recombinant proteins present in crude cellular extract without prior purification. Beyond generating site-specifically modified proteins, this approach could greatly reduce the cost of producing PEGylated polypeptides for therapeutic applications due to the streamlined nature of the process.

Conclusion.

In this work, we have demonstrated that PFTase can be used to introduce aldehyde functionality near the C-terminus of a protein and that the resulting aldehyde-functionalized proteins can then be modified in a plethora of ways via aniline-catalyzed hydrazone or oxime ligation under mild conditions. We show that if the concentration of the aldehyde–functionalized protein is relatively high (>50 μM), hydrazone ligation is more efficient compared with oxime ligation due to its faster kinetics, whereas if the protein concentration is in the low micromolar range, oxime ligation is more advantageous due to its larger equilibrium constant. An important feature of the chemistry reported here is its reversible nature that can be harnessed to permit efficient release of proteins; covalent immobilization using hydrazone ligation of an aldehyde-containing protein can be followed by subsequent oxime formation to release the polypeptide without denaturation. Using synthetically modified alkoxyamines, a variety of new functionality ranging from fluorescent groups to PEG chains can be appended onto proteins. A second key feature of this approach concerns the enzymatic method for aldehyde incorporation. By capitalizing on the selectivity of the enzymatic process, the initial protein functionalization can be performed using unpurified protein substrates. The resulting modified protein can then be captured via hydrazone formation and released via oxime formation to produce a variety of pure, site-specifically modified protein conjugates. Such a streamlined approach for polypeptide modification could be particularly useful for large-scale production of protein conjugates for therapeutic or industrial applications. It should also be noted that the ligation chemistry described herein

and the Cu(I)-catalyzed click reaction are orthogonal. This opens up the possibility of performing multiple modifications on proteins using different bioorthogonal chemistries. Given that CAAX-box sequences can be appended to the C-terminus of almost any protein, the method reported here should be useful for a variety of applications in protein chemistry.

Funding Sources. This work was supported by the National Institutes of Health (GM058842 and GM084152), the University of Minnesota and the Minnesota Supercomputer Institute.

Supporting Information

General. All synthetic reactions were carried out at rt and stirred magnetically unless otherwise noted. TLC was performed on precoated (250 mm) silica gel 60 F-254 plates (Merck). Plates were visualized by staining with $KMnO_4$ or with a hand-held UV lamp. Flash chromatography silica gel (60–200 mesh, 75–250 μm) was obtained from Mallinckrodt Inc. CH_2Cl_2, CH_3CN, and THF were dried by using a Mbraun solvent purification system. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. ^1H NMR spectra were obtained at 300 or 500 MHz; ^{13}C NMR spectra were obtained at 125 MHz; ^{31}P NMR spectra were obtained at 121 MHz. All NMR spectra were acquired on Varian instruments at 25°C. Chemical shifts are reported in ppm and *J* values are in Hz. Fluorescence assay data were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer. Analytical HPLC was performed on a Beckman model 125/166 instrument, equipped with a diode array UV detector, ABI Analytical Spectroflow 980 fluorescence detector, and a Varian C_{18} column (Microsorb-MV, 5 μm, 4.6x250 mm). Preparative HPLC separations were performed by using a Beckman model 127/166 instrument, equipped with a UV detector and a Phenomenex C_{18} column (Luna, 10 μm, 10x250 mm). MS spectra for synthetic reactions were obtained on a Bruker BioTOF II instrument. MS and LC/MS spectra of modified peptides were obtained with an Applied Biosystems/MDS SCIEX QSTAR® Elite Hybrid LC-MS system. Sep-Pak cartridges were purchased from Waters (Milford, MA). Yeast PFTase was prepared as previously described.[1] Protein LC/MS analyses were performed using a Waters Synapt

G2 Quadropole TOF mass spectrometer instrument. MALDI-MS analyses were performed with a Bruker MALDI TOF spectrometer Instrument.

Materials and Methods. Alexa Fluor 488-aminooxy (**6c**) and Texas red hydrazide (**6b**) were from AnaSpec; hydrazide agarose beads were from Thermo Scientific. PEG aminooxy 10,000 MW was from NOF America Corporation. C_{18} Sep-Pak® cartridges were purchased from Waters. Vydac 218TP54 and 218TP1010 columns were used for analytical and preparative RP-HPLC, respectively. All solvents were of HPLC grade. All other reagents were from Sigma Aldrich.

Abbreviations.

AF, Alexa-Fluoro488 aminooxy;

TR, Texas red hydrazide;

FPP, Farnesyl diphosphate;

DTT, dithiothreitol;

ESI-MS, electrospray ionization mass spectrometry;

Far, farnesyl;

RP-HPLC, reversed-phase high-pressure liquid chromatography;

PB, phosphate buffer;

PEG, polyethylene glycol;

DMAP, 4-dimethylaminopyridine;

HOBt, 1-hydroxybenzotriazole;

HTCU, (2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate);

GIP, glucose-dependent insulinotropic peptide;

DIEA, N,N-Diisopropylethylamine;

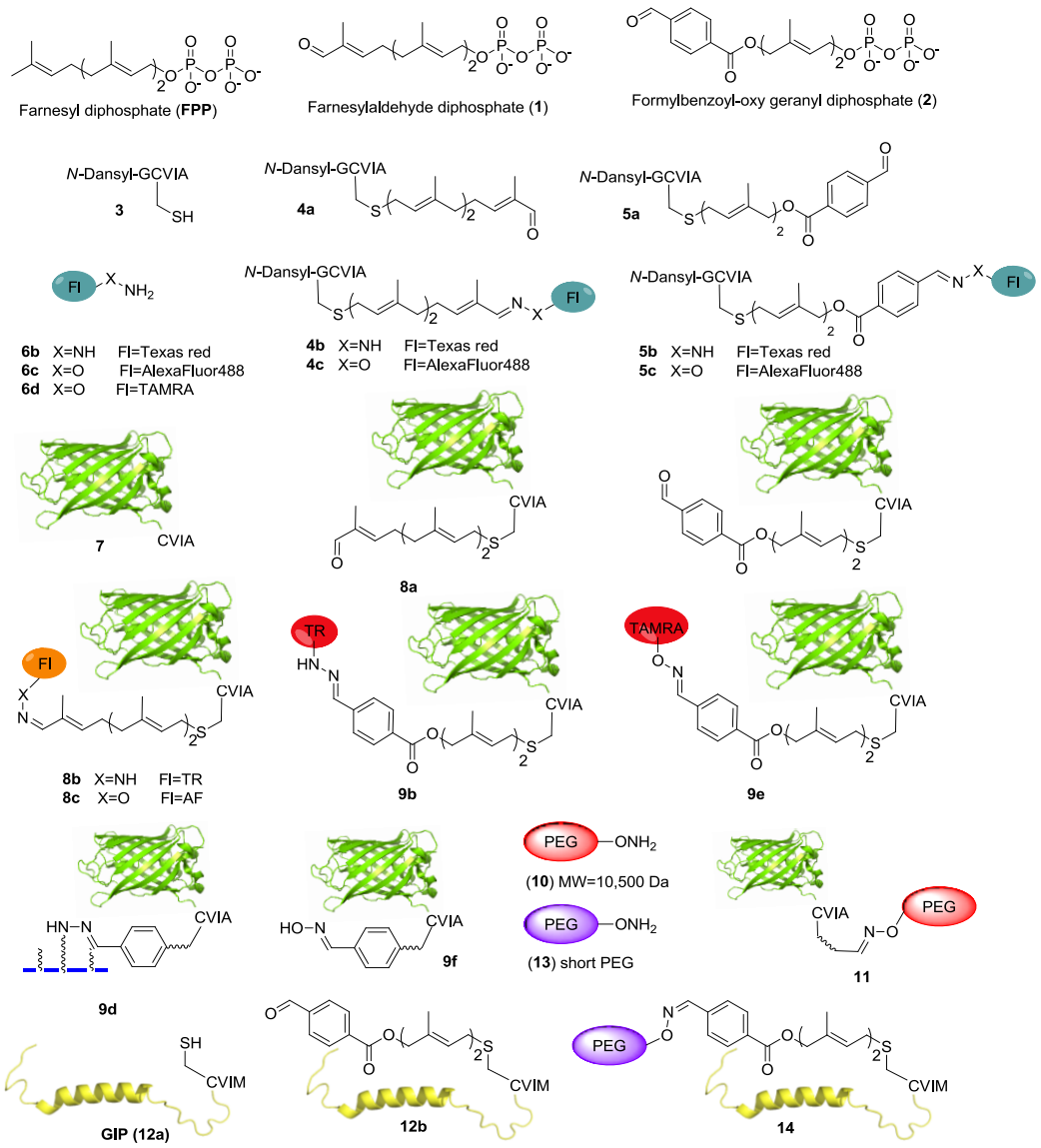
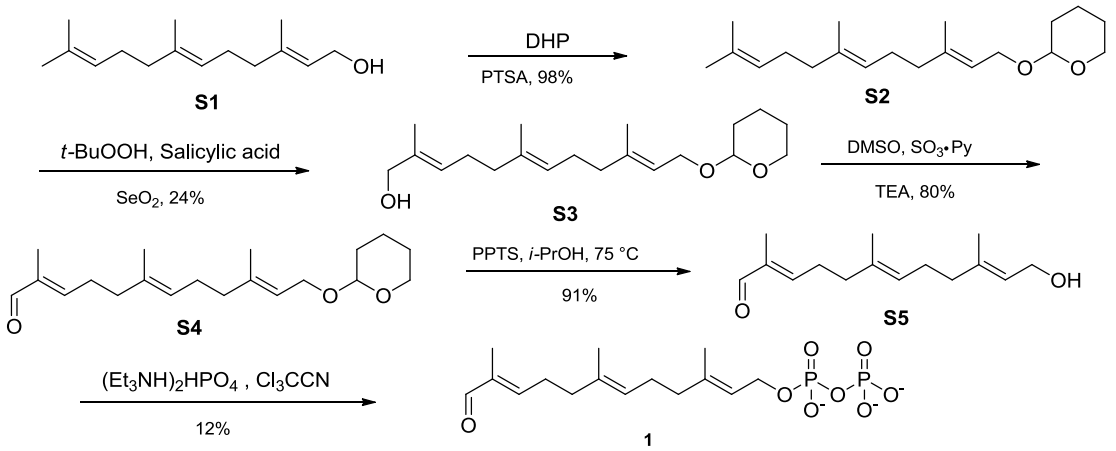


Figure 3.S1. Structures of compounds **1** to **14**.

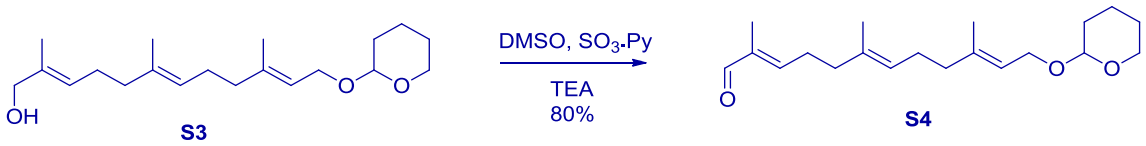
General method used for synthesis of diphosphates from corresponding alcohols. Alcohol (1 eq) was added to CCl_3CN (6 eq) in a 25 mL flask. In a separate 5 mL flask, $(Et_3NH)_2HPO_4$ salt (2.8 eq) was added to dry CH_3CN (1 mL per 50 mg of alcohol in the reaction), and placed in an oil bath at 30 °C for 5 min to dissolve the salt. This solution was added drop-wise to the mixture of alcohol and CCl_3CN solution over 3 h at rt, and was left to stir for an additional 15 min at rt. The slow addition of salt solution to the reaction flask was critical to significantly increase the yield. The solvent was removed *in vacuo* and NH_4HCO_3 (25 mM, 8 mL per 50 mg of alcohol in the reaction) was added to the resulting solution and a white precipitate was formed. The solution was filtered and purified by RP-HPLC with a semi-preparative column under the following conditions: detection at 214 nm; flow rate at 5.0 mL·min^{-1}; 5 mL injection loop; solvent A: 25 mM NH_4HCO_3 in H_2O, solvent B: CH_3CN. The specific HPLC method used and the yield obtained for each substrate is described in its respective section**.**



Scheme 3.S1. Synthesis of farnesyl aldehyde diphosphate (**1**).

Synthesis of farnesyl aldehyde diphosphate (1)

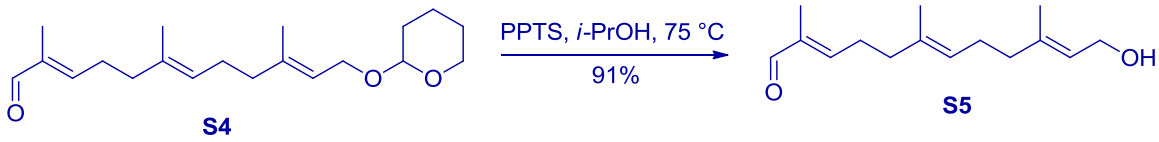
(2E,6E,10E)-2,6,10-trimethyl-12-*O*-THP-dodeca-2,6,10-trienal (S4).



Compounds **S2** and **S3** were prepared as previously described.[2] Alcohol **S3** (1.00 g, 3.10 mmol) was dissolved in anhydrous CH_2Cl_2 (20 mL) in a 50 mL flask and the solution was cooled to 0 °C in an ice bath. DMSO (3.1 mL, 43 mmol) was added drop-wise to the solution mixture followed by the addition of triethylamine (2.75 mL, 197 mmol). Sulfur trioxide pyridine complex (SO_3·Py, 2.52 g, 15.8 mmol) was added slowly over 10 min to the reaction mixture. The reaction was stirred at 0 °C for an additional 1 h until TLC analysis showed almost complete conversion to the product. The reaction was stopped by the addition of 100 mL CH_2Cl_2 and washed with 5 M HCl (2 x 10 mL) until the aqueous phase remained acidic (tested via pH paper) indicating the absence of base in the reaction mixture. Next, the solution was washed with sat. $NaHCO_3$ (1 x 15 mL) followed by brine (2 x 10 mL). The solution was dried over Na_2SO_4 and the solvent was removed *in vacuo*. Crude product was purified by silica gel column chromatography using a step gradient of solvent (Hex:EtOAc) starting from 1:0 (v/v) going to 3:1 (v/v) to afford 0.79 g of compound **S4** (2.48 mmol) as a pale yellow oil (80% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.37 (s, 1H), 6.46 (t, J = 7.2 Hz, 1H), 6.46 (t, J = 7.2 Hz, 1H), 5.36 (dd, J = 6.3 Hz, J = 7.3 Hz, 1H), 5.16 (t, J = 6.2 Hz, 3 H), 4.62 (dd, J = 3.0 Hz, J = 4.5 Hz, 1H), 4.24 (dd, J =

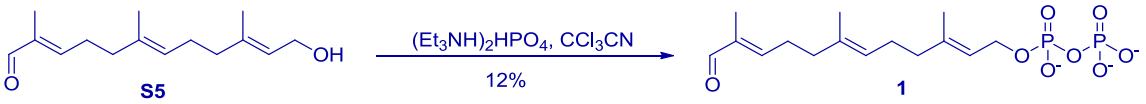
6.3 Hz, *J* = 12 Hz, 1H), 4.02 (dd, *J* = 7.3 Hz, *J* = 12 Hz, 1H), 3.51 (m, 1H), 3.89 (m, 1H), 2.0-2.25 (m, 6H), 2.46 (m, 2H), 1.74 (s, 3H), 1.82 (m, 1H), 1.63 (s, 3H), 1.67 (s, 3H), 1.55 (m, 5H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 195.38, 154.53, 139.95, 133.71, 125.33, 121.75, 120.86, 97.91, 63.70, 62.35, 39.51, 38.01, 30.77, 27.45, 26.25, 25.55, 19.69, 16.47, 15.97, 9.28. HR-ESI-MS calcd for $C_{20}H_{32}O_3Na$ $[M+Na]^+$ 343.2249, found 343.2246.

(2E,6E,10E)-12-hydroxy-2,6,10-trimethyldodeca-2,6,10-trienal (S5).

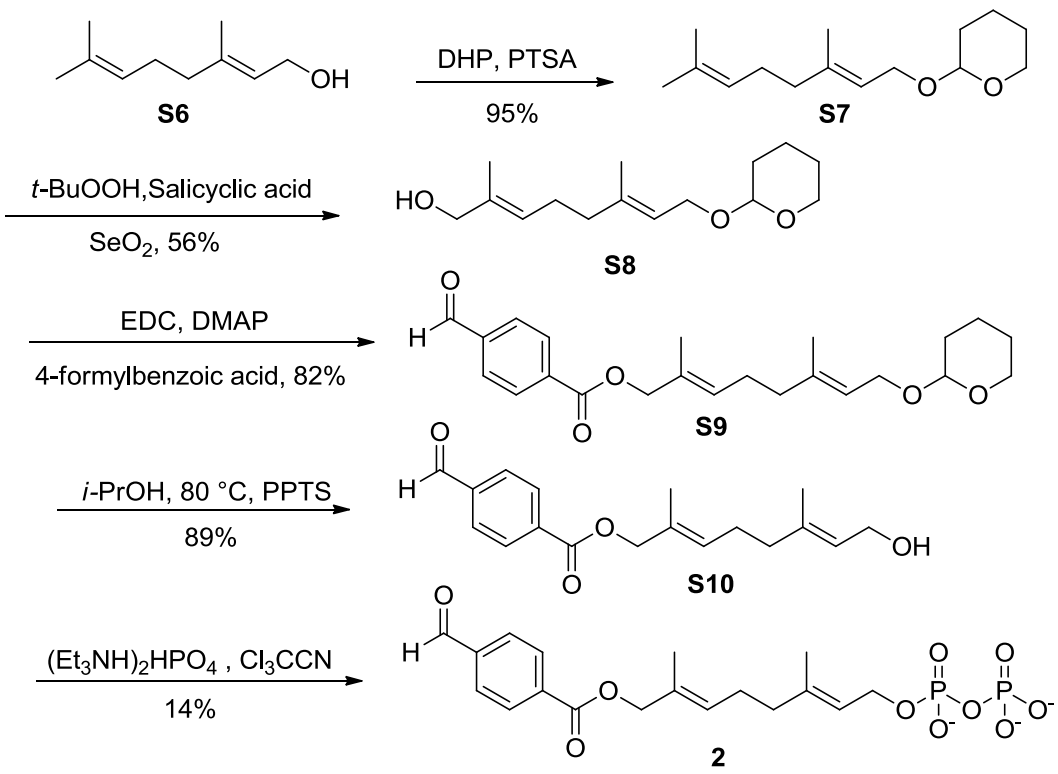


Protected aldehyde **S4** (1.30 g, 4.05 mmol) was dissolved in *i*-PrOH (20 mL) in a 50 mL flask. PPTS (30 mg) was added as catalyst. The reaction was then refluxed at 75 °C for 4 h, when TLC analysis indicated complete conversion to the product. It was then quenched by adding sat. $NaHCO_3$ (10 mL) and EtOAc (100 mL). The organic layer was then separated and dried over Na_2SO_4. The solvent was removed *in vacuo* and afforded 0.87 g of compound **S5** (91% yield) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.37 (s, 1H), 6.48 (t, *J* = 7.5 Hz, 1H), 5.41 (t, *J* = 4.5), 5.16 (t, *J* = 7 Hz, 3 H), 4.14 (t, *J* = 7 Hz, 1H), 2.45 (t, *J* = 7.5, 2H), 2.0-2.25 (m, 6H), 1.68 (s, 3H), 1.74 (s, 3H), 1.64 (s, 3H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 195.35, 154.62, 139.26, 133.60, 125.16, 123.72, 117.30, 63.57, 39.33, 38.08, 27.34, 26.17, 16.20, 15.85, 9.16. HR-ESI-MS calcd for $C_{15}H_{24}O_2Na$ $[M+Na]^+$ 259.1674, found 259.1669.

(2E,6E,10E)-3,7,11-trimethyl-12-oxododeca-2,6,10-trien-1-yl diphosphate (1).



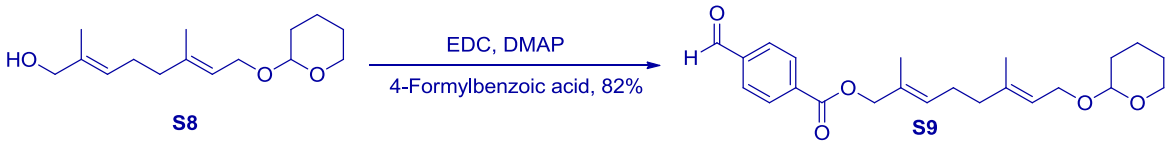
Aldehyde **S5** (0.18 g, 0.76 mmol) was used to synthesize compound **1** using the "General Method used for Synthesis of diphosphates." The reagents used were CCl_3CN (457 µL, 4.56 mmol), $(Et_3NH)_2HPO_4$ (0.57 g, 1.90 mmol) and CH_3CN (3.6 mL). The RP-HPLC method used was a gradient of 0–30% solvent B in 30 min, 30–100% in 5 min; compound **1** eluted from 20–25% solvent B. Fractions containing pure **1** were collected and lyophilized. The resulting salt was dissolved in D_2O and its concentration was measured following a previously established NMR-based quantification[3] to yield 10 mL of 2.5 mM solution of **1** (36 mg, 12% yield). ^{1}H NMR: (500 MHz, D_2O) δ 9.11 (s, 1H), 6.61 (t, J = 7.5 Hz, 1H), 5.29 (t, J = 7 Hz, 1H), 5.08 (t, J = 6 Hz, 1H), 4.30 (t, J = 6 Hz, 2H), 2.38 (t, J = 7 Hz, 2H), 2.07 (t, J = 7 Hz, 2H), 2.00 (t, J = 7 Hz, 2H), 1.917 (t, J = 7 Hz, 2H), 1.54 (s, 6H), 1.49 (s, 3H). ^{31}P NMR: (121 MHz, D_2O) δ -5.971 (d, J = 22.6, 1P), -10.013 (d, J = 22.6, 1P). HR-ESI-MS calcd for $C_{15}H_{26}O_8P_2$ $[M\text{-}H]^-$ 395.1025, found 395.0907.



Scheme 3.S2. Synthesis of formyl benzoyl-oxy geranyl diphosphate (**2**).

Synthesis of formyl benzoyl-oxy geranyl diphosphate (2).

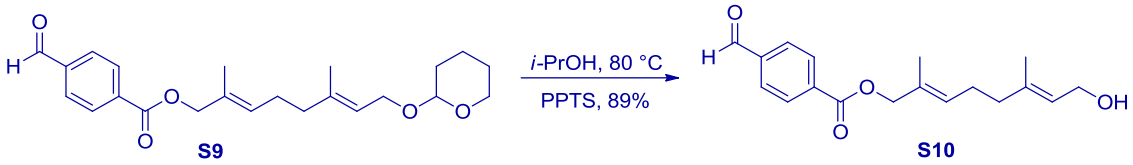
(2E,6E)-2,6-dimethyl-8-O-THP-octa-2,6-dien-1-yl 4-formylbenzoate (S9).



Compounds **S7** and **S8** were prepared as previously described.[2] Alcohol **S8** (1.50 g, 5.90 mmol) was dissolved in CH_2Cl_2 (20 mL) in a flame dried 50 mL flask. *para*-Formylbenzoic acid (0.73 g, 7.87 mmol) and DMAP (0.43 g, 5.31 mmol) were added to the reaction flask, then left to stir for 10 min at 0 °C. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) (1.83 g, 11.8 mmol) was then added to the reaction flask, and stirred

at 0 °C for 1 h until TLC analysis (2:1 Hex:EtOAc v/v) indicated almost complete conversion to the product. The reaction was diluted in EtOAc (100 mL), and treated with 5% HCl (1:10 v/v) until the aqueous layer remained acidic (tested via pH paper). The reaction was then washed with sat. $NaHCO_3$ (2 x 10 mL) and brine (2 x 5 mL). The solvent was removed *in vacuo* and the crude product was further purified by silica gel column chromatography with gradient elution (Hex:EtOAc) from 10:1 (v/v) going to 2:1 (v/v) to afford 1.86 g of compound **S9** (4.84 mmol, 82% yield) as a pale yellow oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 10.11 (s, 1H), 8.21 (d, J = 8.0 Hz, 2H), 7.96 (d, J = 8.4 Hz, 2H), 5.57 (t, J = 7.9 Hz, 1H), 5.39 (t, J = 7.3 Hz, 1H), 4.74 (s, 2H), 4.62 (m, 1H), 4.25 (dd, J = 11.7, 6.4 Hz, 2H), 4.03 (dd, J = 11.8, 7.4 Hz, 1H), 3.95–3.83 (m, 1H), 3.51 (dd, J = 10.7, 4.8 Hz, 1H), 2.24 (m, 2H), 2.10 (t, J = 7.3 Hz, 2H), 1.76 (s, 3H), 1.70 (s, 3H), 1.66–1.46 (m, 5H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 191.50, 165.24, 139.32, 138.97, 135.23, 130.06, 129.81, 129.65, 129.38, 120.98, 97.75, 96.82, 71.11, 63.49, 62.18, 38.79, 30.58, 25.90, 25.36, 19.51, 16.31, 13.98.

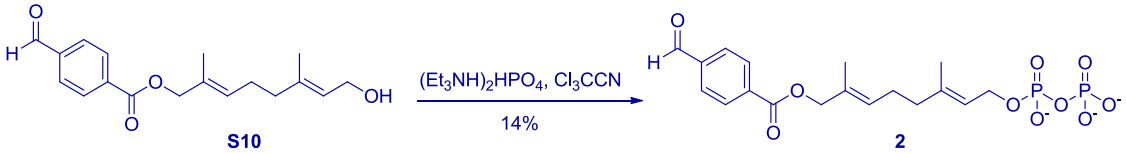
(2E,6E)-8-hydroxy-2,6-dimethylocta-2,6-dien-1-yl 4-formylbenzoate (S10).



Protected alcohol **S9** (1.26 g, 3.26 mmol) was dissolved in *i*-PrOH (20 mL) in a 50 mL flask. PPTS (30 mg) was added as a catalyst. The reaction was then refluxed at 80 °C for 4 h when TLC analysis (2:1 Hex:EtOAc v/v) showed almost complete conversion to the product, and was quenched by adding sat. $NaHCO_3$ (10 mL) and EtOAc (100 mL). The

organic layer was separated and dried over Na_2SO_4. Solvent was removed *in vacuo* to afford 0.87 g of compound **S10** (2.90 mmol, 89% yield) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 10.10 (s, 1H), 8.21 (d, J = 8.1 Hz, 2H), 7.96 (d, J = 8.1 Hz, 2H), 5.56 (t, J = 6.7 Hz, 1H), 5.43 (t, J = 6.5 Hz, 1H), 4.75 (s, 2H), 4.16 (d, J = 6.8 Hz, 2H), 2.23 (m, 2H), 2.10 (t, J = 7.5 Hz, 2H), 1.76 (s, 3H), 1.68 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 191.71, 165.43, 139.09, 138.44, 135.30, 130.17, 129.91, 129.61, 129.52, 124.06, 71.21, 59.18, 38.79, 25.91, 16.19, 14.06. HR-ESI-MS calcd for $C_{18}H_{22}O_4Na$ $[M+Na]^+$ 325.1415, found 325.1477.

1-(((2E,6E)-8-((4-formylbenzoyl)oxy)-3,7-dimethylocta-2,6-dien-1-yl)oxy)-1,3-dihydroxy-3-oxido-1,3-dioxodiphosphoxane (2).



Aldehyde **S10** (0.52 g, 1.72 mmol) was used to synthesize compound **2** using the "General Method used for Synthesis of diphosphates." The reagents used were CCl_3CN (1.035 mL, 10.32 mmol), $(Et_3NH)_2HPO_4$ (1.55 g, 5.16 mmol) and CH_3CN (4 mL). The RP-HPLC method used was a gradient of 0–35% solvent B in 30 min, 35–100% in 5 min; compound **2** eluted from 25–30% solvent B . Fractions containing pure **2** were collected and lyophilized. The resulting salt was dissolved in D_2O and concentration was measured following the previously established NMR-based quantification method[3] providing 20 mL of 8.72 mM of compound **2** (80 mg, 15% yield). The product was stored in Tris·HCl (50 mM, pH 7.5) at -80°C; ^{1}H NMR (500 MHz, D_2O): δ 9.81 (s, 1H), 7.90 (d, J = 6.5 Hz, 2H), 7.78 (d, J = 7.2 Hz, 2H), 5.39 (t, J = 6.5 Hz, 1H), 5.27 (t, J = 6.5 Hz, 1H), 4.53 (s,

2H), 4.29 (t, J = 6.6 Hz, 2H), 2.04 (m, 2H), 1.94 (t, J = 6.9 Hz, 2H), 1.53 (s, 6H). ^{31}P NMR (162 MHz, D_2O) δ -5.91 (d, J = 22.4 Hz), -9.73 (d, J = 22.4 Hz). HR-ESI-MS calcd for $C_{18}H_{23}O_{10}P_2$ $[M-H]^-$ 461.0771, found 461.2021.

Enzymatic studies of FPP-analogues 1 and 2 using a continuous fluorescence assay. Enzymatic reaction mixtures contained Tris·HCl (50 mM , pH 7.5) , $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), 2.4 μM *N*-dansyl-GCVIA (**3**), 0.040 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, PFTase (80 nM), and varying concentrations of either **1** or **2** (0-50 μM), in a final volume of 250 μL. The reaction mixtures were equilibrated at 30 °C for 5 min, initiated by the addition of PFTase, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 10 min. The initial rates of formation of products were obtained as slopes in IU/min using least squares analysis. Corrections were applied to all the rate calculations based on the difference between the fluorescence intensity of the prenylated product and the starting peptide. Assuming 100% conversion, the difference corresponds only to the fluorescence of the total amount of the product. The slope was then divided by the fluorescence difference followed by multiplying by the total concentration of peptide (2.4 μM) which then gives the rate of formation of product in μM/s. It should be noted that the K_M values reported here are actually apparent K_M values,

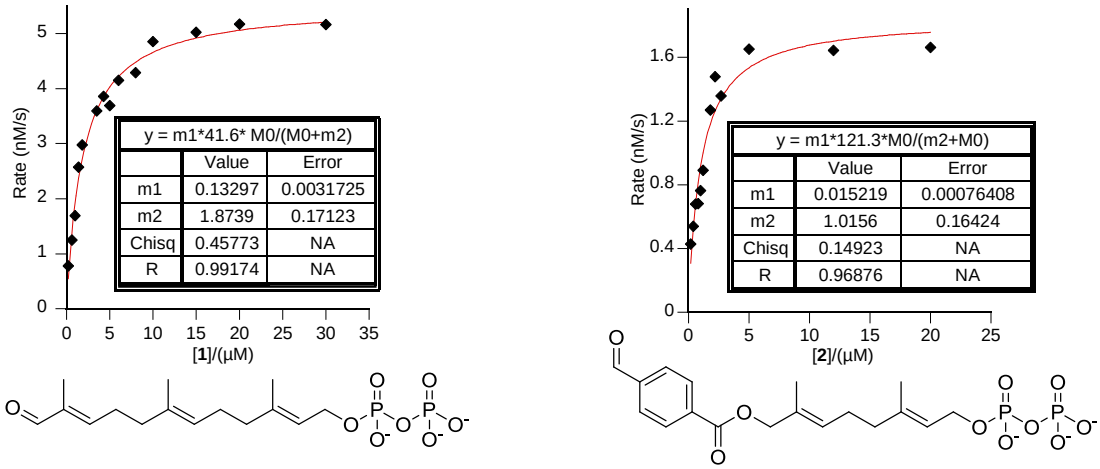


Figure 3.S2. Fluorescence-based PFTase enzyme assay for prenylation of model peptide **3** using varying concentrations of **1** (left panel) or **2** (right panel).

since the measurements were performed in only a single peptide concentration. The data were fit to a Michaelis-Menten model ($V = \frac{[\mathrm{E}_\circ]k_{cat}*[\mathrm{S}]}{K_M+[S]}$) using a nonlinear regression program, to determine k_{cat} and K_M.

Oxime ligation between peptide-aldehyde 4a and 5a and aminooxy alexafluor-488 (6c). Coupling reactions contained 3-5 μM **4a** or **5a**, 200 μM alexafluor-488 (**6c**), PB (0.1 M, pH 7.0), and aniline (100 mM) in a final volume of 500 μL. Reactions were performed at rt and were initiated by addition of aniline (100 mM). LC-MS analysis of the reaction mixture after 3-4 h gave ions of 1384.6 Da and 1450.6 Da as the predominant species, which are consistent with $[M+H]^+$ for **4c** and **5c,** respectively.

Hydrazone ligation between peptide-aldehydes 4a and 5a and Texas red hydrazide (6b). Coupling reactions contained 3-5 μM **4a** or **5a**, 200 μM Texas red (**6b**), PB (0.1 M, pH 7.0), and aniline (100 mM) in final volume of 500 μL. Reactions were performed at rt

and were initiated by addition of aniline (100 mM). LC-MS analysis of the reaction mixture after 1 h gave ions of 758.46 Da and 791.45 Da as the predominant species, consistent with $[M+2H]^{2+}$ for hydrazones **4b** and **5b**, respectively.

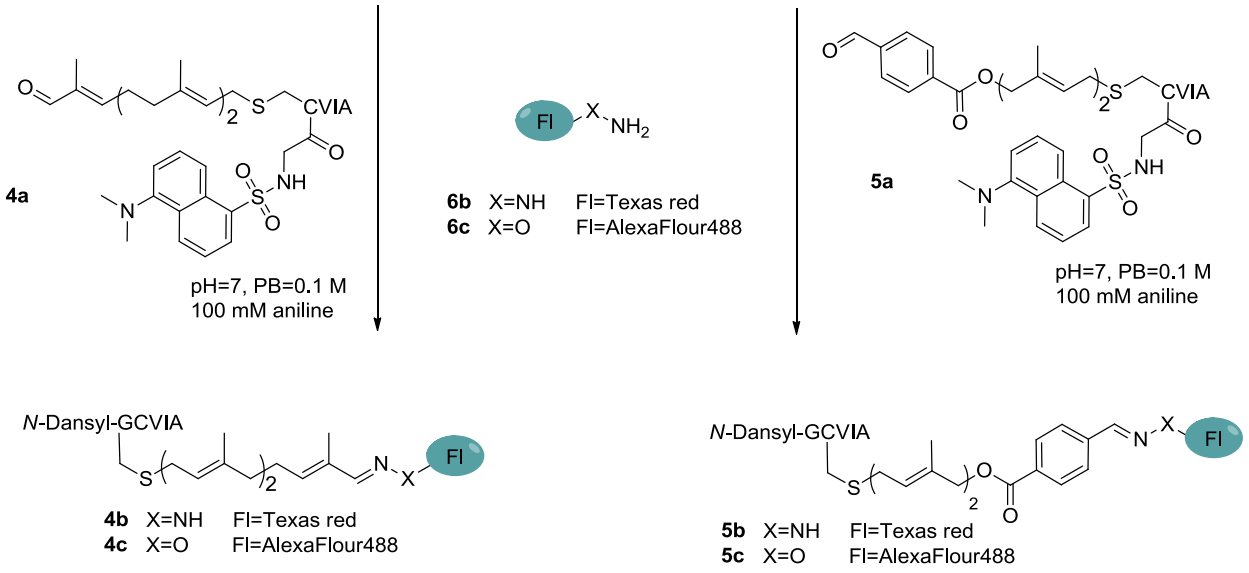


Figure 3.S3. Schematic representation of oxime and hydrazone ligation reactions of aldehydes **4a** and **5a** with **6b** and **6c** to form **4b** and **4c,** and **5b** and **5c** respectively.

GFP-CVIA. Protein was prepared as previously described with one modification.[4,5] In the final phenyl sepharose chromatography step, after the protein was loaded onto the column and washed with buffer as explained in the original work, the protein was eluted from column by adding water instead of buffer.

LC-MS analysis of GFP for determination of prenylation efficiency. Purified prenylated GFP was analyzed by LC-MS to ensure complete prenylation. The only peak observed was prenylated GFP in case of **1**, and there was a small trace of remaining unprenylated GFP in case of **2**. Proteins were stored in Tris·HCl (50 mM, pH 7.5) prior to

injection into the LC-MS instrument. The LC-MS method used was gradient 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

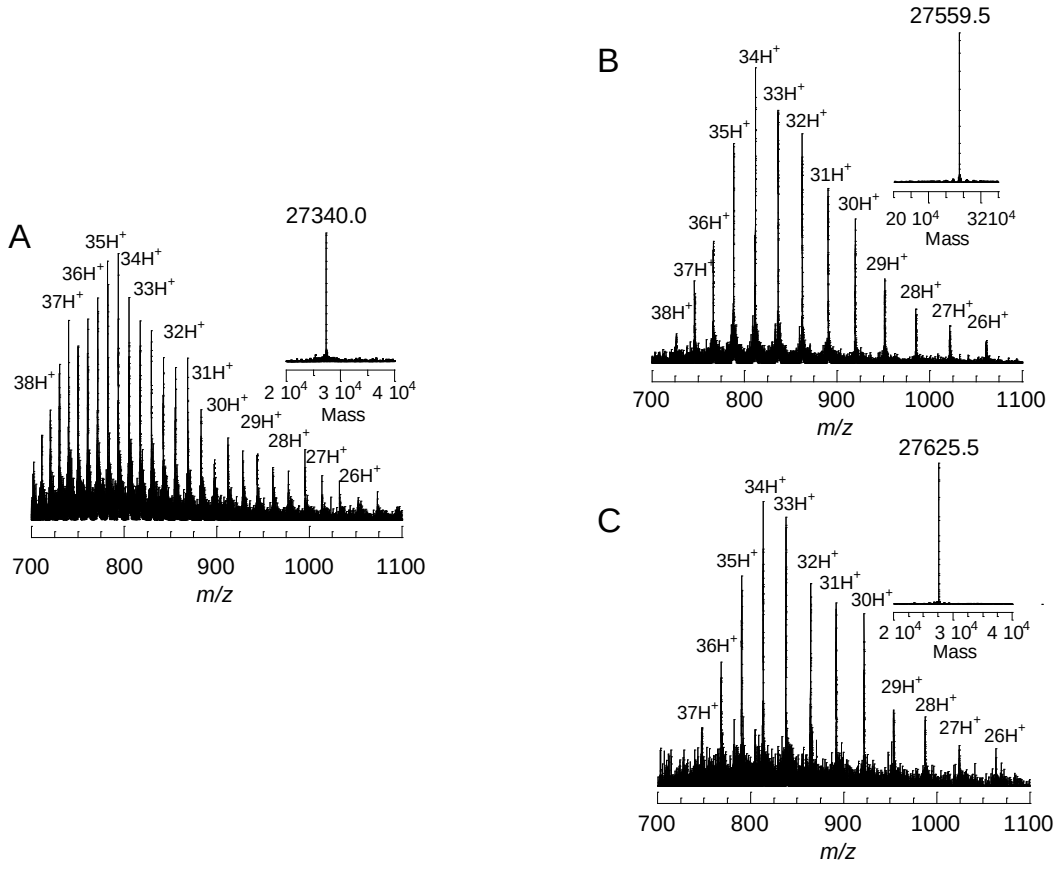


Figure 3.S4. ESI mass spectra of A) GFP-CVIA (**7**), B) GFP prenylated with aldehyde-analogue **1** to yield **8a,** and C) GFP prenylated with aldehyde-analogue **2** to yield **9a**, showing succesful prenylation of GFP in both cases. The deconvoluted mass spectra are shown in the insets.

Effect of prenylation on GFP fluorescence. Fluorescence of same concentration (9 µM) of GFP-CVIA (**7**) and GFP-aldehyde (**9a**) were measured and the result showed that prenylation does not affect the protein fluorescence characteristic.

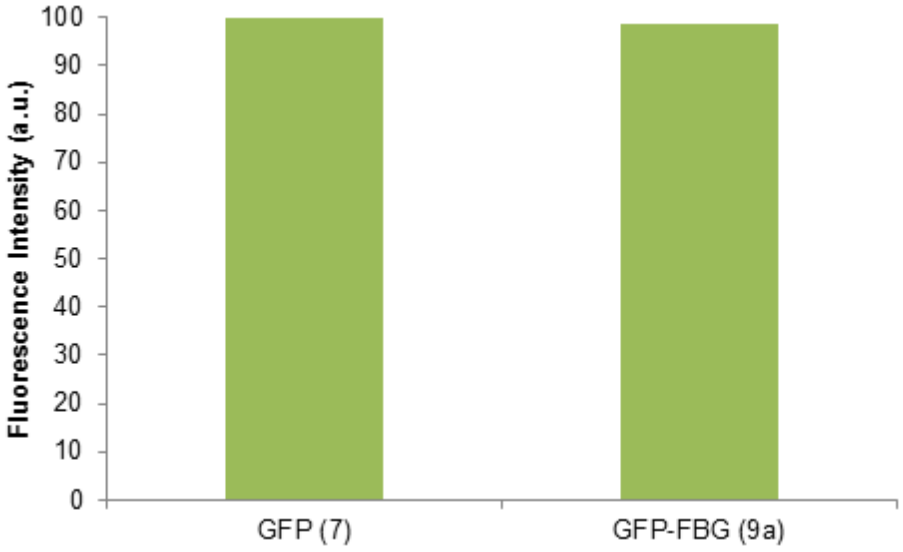


Figure 3.S5. Comparison of the fluorescence intensity of GFP-CVIA (**7**) and aldehyde-GFP (**9a**). The fluorescence of solutions containing **7** or **9a** (9 μM) were measured to establish that prenylation does not affect the fluorescence properties of GFP (excitation at 450 nm and emission at 535 nm).

Molecular modeling of GFP-TAMRA. The model of GFP-TAMRA was based on the previously solved X-ray structure of GFP (PDB ID: 2Y0G) and subsequent modeling was performed using Maestro (Schrodinger, Portland, OR). The CVIA tag was first appended onto the C-terminus of GFP and the prenyl moiety containing the covalently coupled TAMRA was built onto the thiol side chain of the C terminal cysteine. The protein, the CVIA C-terminus and the prenyl-fluorophore group were then energy minimized using a truncated Newton algorithm.

Immobilization of 9a onto hydrazide agarose beads. Hydrazide agarose beads (Thermo Scientific, hydrazide loading: 16 µmol/mL) (300 µL) were washed with PB (0.1 M, 3x500 µL, pH 7.0). PB (30 µL, 1 M, pH 7.0) was added to the beads followed by addition of **9a** (200 µL, 87 µM). Immobilization was initiated by adding aniline (2 µL, 100 mM). For controls, GFP-CVIA (**7**) was added instead of **9a**. The solution was centrifuged, then the GFP UV-absorbance of the supernatant was measured as a function of time (488 nm, ε=55,000 M^{-1} ·cm^{-1}). After 2 h, the solution was centrifuged and the beads were washed thoroughly with PB (0.3 M, pH 7.3, 3x300 µL) and KCl (3x300 µL, 1 M) to remove non-specifically bound proteins and were stored in pH 7.5 Tris buffer at 4 °C.

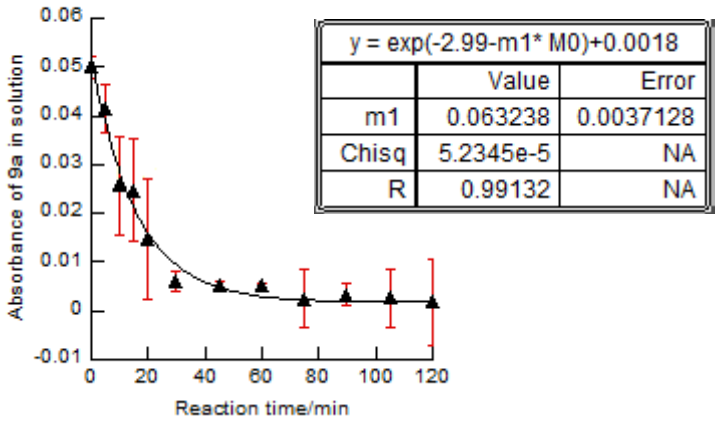


Figure 3.S6. Kinetic analysis of immobilization of **9a** onto hydrazide functionalized agarose beads. The reaction was carried out at rt, in the presence of 100 mM aniline and excess beads. UV absorbance of GFP in the supernatant was measured at different times showing >95% immobilization in ~45 min. Experiments were performed in duplicate. The data was fit to a simple exponential process.

MALDI analysis of protein PEGylation. MALDI analysis was used to measure the M_n, M_w and PDI of the PEG **10** and GFP-PEG **11** using equations below:

$$M_n = \frac{\sum M_i N_i}{\sum N_i} \quad ; \quad M_w = \frac{\sum M_i^2 * N_i}{\sum N_i} \quad ; \quad PDI = \frac{M_w}{M_n}$$

Where M is the mass of each species and N is the intensity of that specific peak. Calculations showed that for PEG **10**: M_n=10,790, M_w=10,804 and PDI=1.0012. MALDI analysis for GFP-PEG **11** showed M_n=38,637, M_w=37,647 and PDI=1.0003. It should be noted that since PDI is, in part, related to the mass of the polymer, it is expected that addition of a monodisperse protein to a polydisperse polymer should result in a decrease in PDI of the protein-polymer conjugate as was observed.

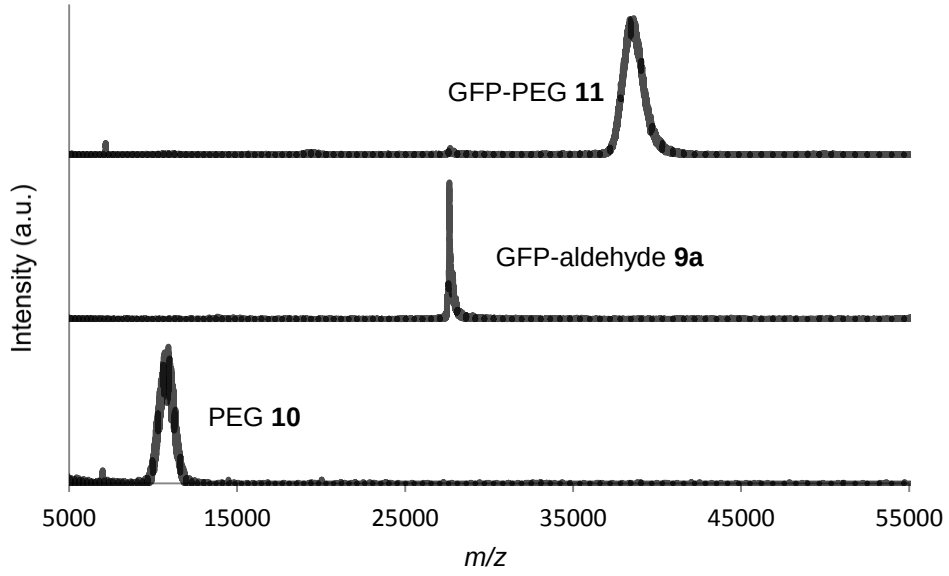
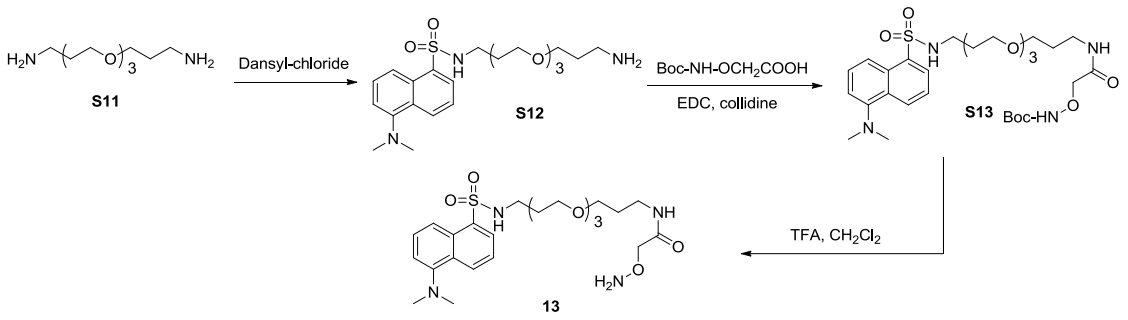


Figure 3.S7. MALDI analysis of PEG **10**, GFP-aldehyde **9a** and PEGylated GFP **11**. The lower panel is the MALDI spectrum of pure PEG **10**, the middle panel is the MALDI spectrum of pure **9a** and the top panel is the MALDI spectrum of the oxime PEGylated GFP **11**, which confirms complete conversion. Excess of **10** was removed via a zip-tip protocol prior to MALDI analysis of **11**.



Scheme 3.3. Synthesis of PEG-aminooxy (**13**).

Synthesis of PEG 13. 4,7,10-Trioxa-1,13-tridecanediamine (**S11**, 9.3 mmol, 2.049 g) was dissolved in anhydrous DMF (15 mL) in a 25 mL round bottom flask. Dansyl-chloride (1.86 mmol, 500 mg) was added to the solution and the reaction mixture was stirred at rt for 3 h. The reaction was stopped by the addition of 100 mL CH_2Cl_2 and washed with 5 M NaOH (2 x 10 mL). Next, the solution was washed with brine (1 x 10 mL) and was dried over Na_2SO_4. The solvent was removed *in vacuo*, the crude product was re-dissolved in 4 mL DMF, and then one fourth of the resulting solution was transferred to a 2 mL eppendorf tube. (Boc-aminooxy)acetic acid (0.5 mmol, 95 mg) was dissolved in 0.5 mL DMF. Collidine (1.1 mmol, 145 µL) and EDC (0.5 mmol, 88.5 µL) were added to the (Boc-aminooxy)acetic acid solution and the mixture was quickly added to the PEG containing solution. The reaction mixture was stirred for 2 h at rt. The reaction was stopped by the addition of 100 mL EtOAc and washed with 5 M NaOH (1 x 5 mL), sat. NH_4Cl (1 x 5 mL) and 5 M HCl (1 x 5 mL) respectively. Next, the solution was washed with brine (1 x 5 mL) and was dried over Na_2SO_4. The solvent was removed *in vacuo*, the crude product was dissolved in 1 mL CH_2Cl_2, and transferred to a 5 mL flask. CH_2Cl_2 (2 mL) and TFA (1.5 mL) were added to the crude product and the reaction

mixture was stirred for 2 h at rt. Next the solvent was removed *in vacuo* and the crude product was dissolved in 5 mL solution of H_2O containing 0.1% TFA (v/v). The product PEG **13** was purified by RP-HPLC with a semi-preparative column under the following conditions: detection at 330 nm; flow rate at 5.0 mL·min^{-1}; 1 mL injection loop; solvent A: H_2O, solvent B: CH_3CN. The RP-HPLC method used was a gradient of 0–35% solvent B in 30 min, 35–100% in 5 min; PEG **13** eluted from 23–28% solvent B. Fractions containing pure **13** were collected and lyophilized. HR-ESI-MS calcd for $C_{24}H_{39}O_7N_4S$ $[M+H]^+$ 527.2534, found 527.2571. The compound was dissolved in 2 mL Tris·HCl (50 mM, pH 7.5) and stored at -20 °C. The concentration was measured using UV absorbance of dansyl-chromophore at 325 nm (ε=4200 M^{-1}·cm^{-1}, 4.3 mM, 4.5 mg, 2% yield over three steps).

Synthesis of GIP-CVIM. GIP-CVIM is a fusion of GIP[6] and the 15 C-terminal residues from human K-Ras 4B.[7] The sequence is shown below in Figure 3.S8. The protein was prepared by solid phase synthesis using a Protein Technologies PS3 synthesizer employing standard Fmoc procedures. The synthesis was performed on a 0.1 mmol scale starting with preloaded Fmoc-L-Met-PEG-PS resin (Applied Biosystems). Each residue was double-coupled in the presence of HCTU, HOBT and DIEA. Following completion of the synthesis, the material was treated with Reagent K for 2.5 h followed by precipitation with Et_2O. The solid product was isolated by centrifugation and the pellet washed twice with Et_2O. The resulting white solid was dissolved in H_2O containing 0.1% TFA and half of it was purified by RP-HPLC. Yield 43.1 mg (13% from 50 μmol peptide

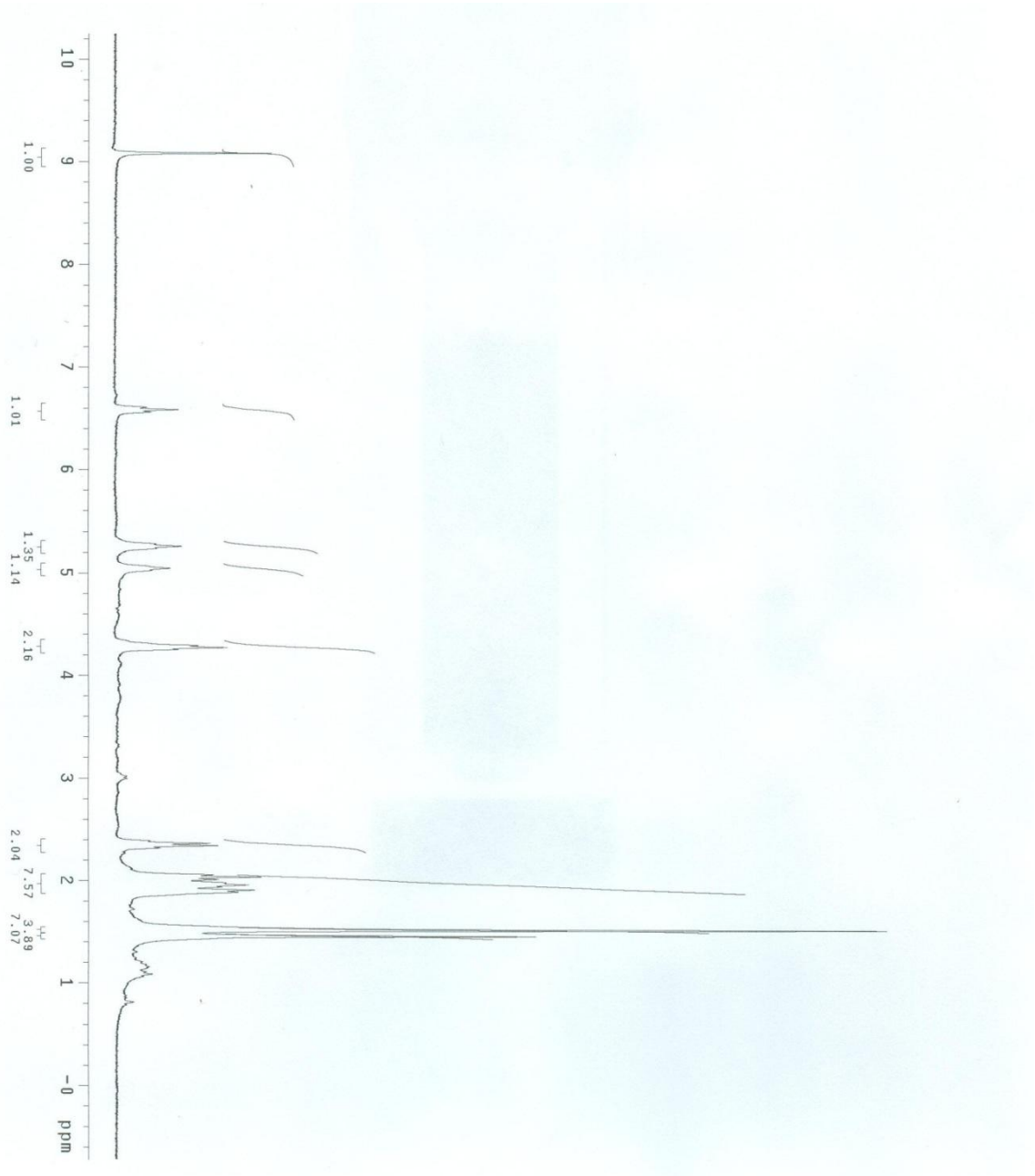
resin), purity by LC-MS analysis: 83%. LC-ESI-MS: calcd: 6700.8, observed (deconvoluted): 6700.0.

YAEGTFISDYSIAMDKIHQQDFVNWLLAQKGKKNDWKHNITQGKKKKKKSKTK

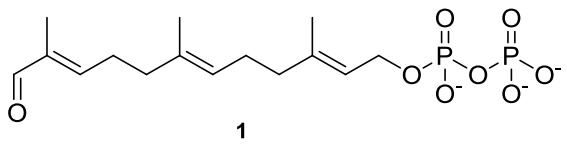
CVIM

Figure 3.S8. Sequence of GIP-CVIM. The 15 C-terminal residues from human K-Ras 4B are underlined.

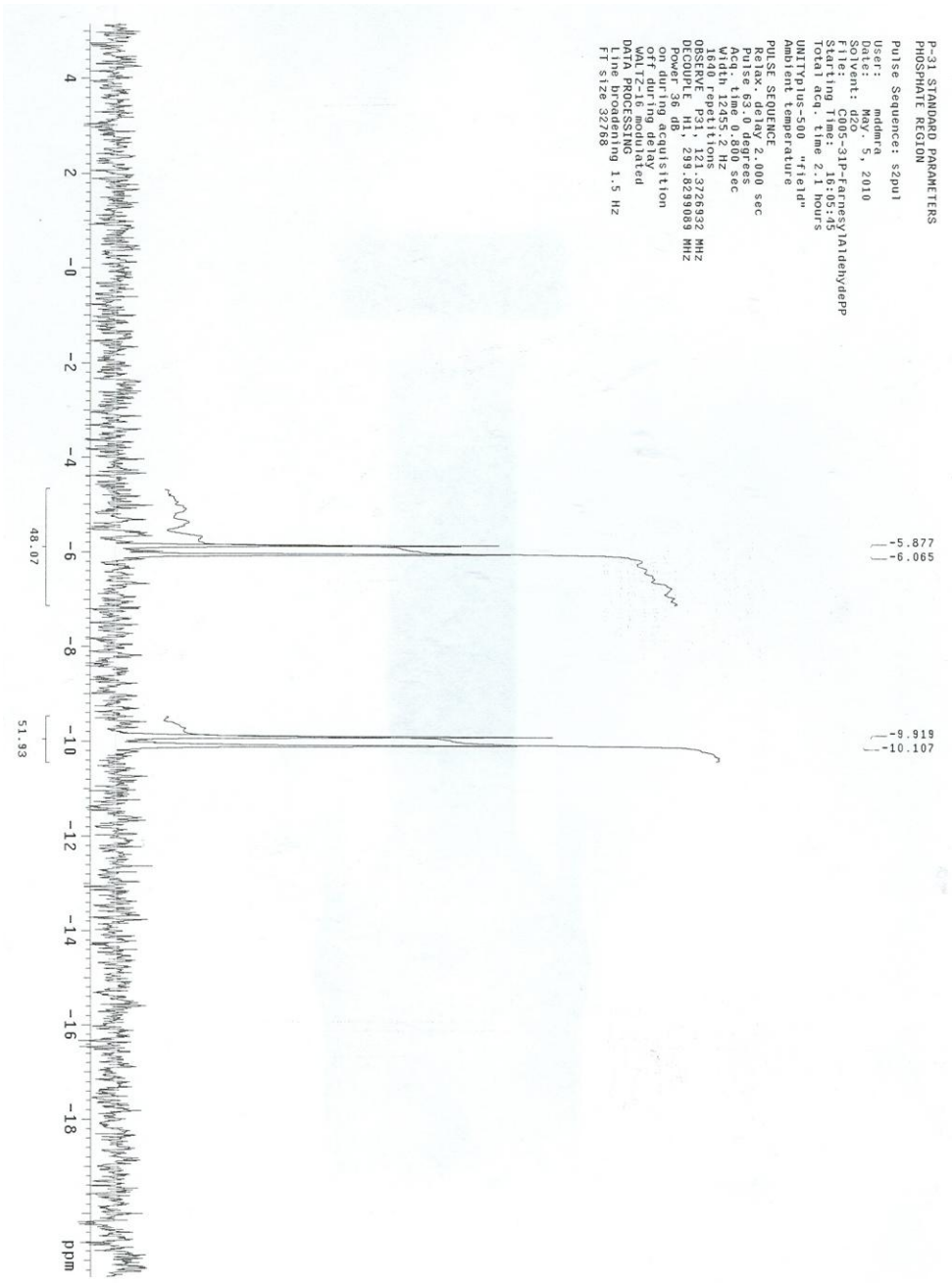
NMRs



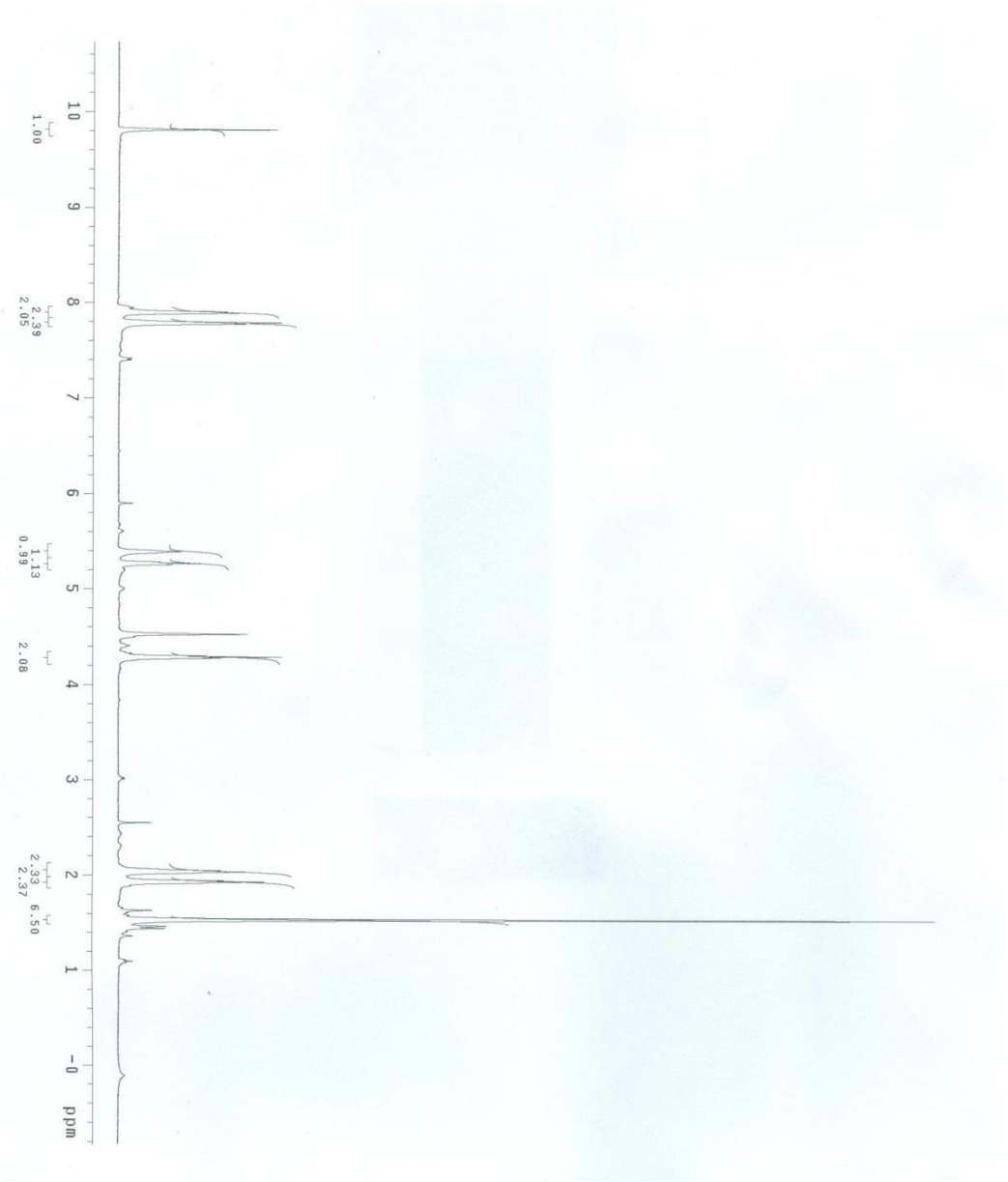
^{1}H-NMR for compound **1**.



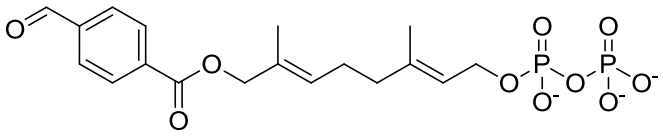
1

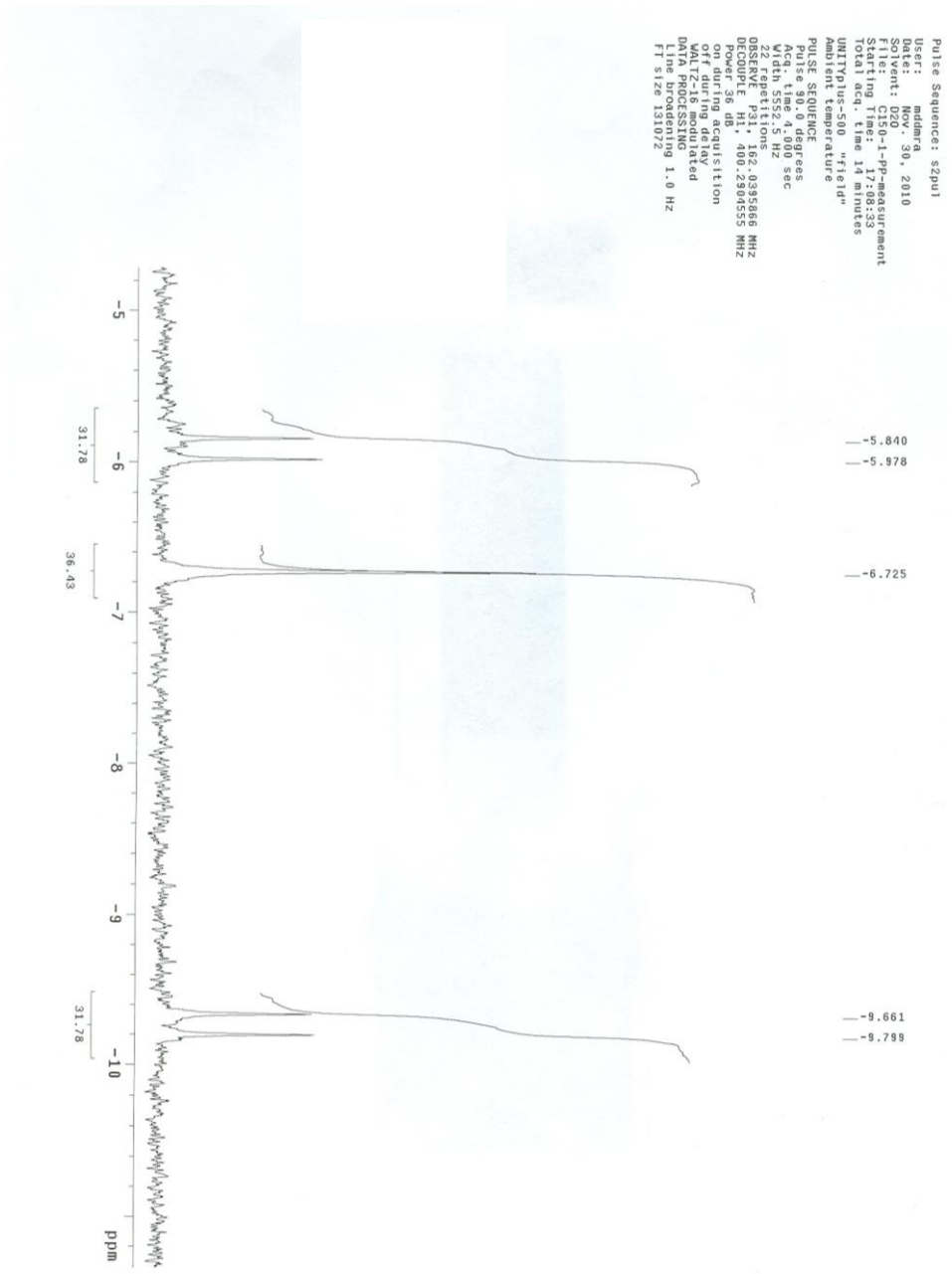


^{31}P-NMR for compound **1**.



^{1}H-NMR of compound **2**.





^{31}P-NMR of compound **2**. $Na_2H_2P_2O_7$ is used as standard for quantification measurements.[3]

Chapter 4. A Highly Efficient Catalyst for Oxime Ligation and Hydrazone-Oxime Exchange Suitable for Bioconjugation

Mohammad Rashidian, Mohammad Mohsen Mahmoodi, Mark D. Distefano

Imine-based reactions are useful for a wide range of bioconjugation applications. Although aniline is known to catalyze the oxime ligation reaction under physiological conditions, it suffers from slow reaction kinetics, specifically when a ketone is being used or when hydrazone-oxime exchange is performed. Here, we report on the discovery of a new catalyst that is up to 15 times more efficient than aniline. That catalyst, *m*-phenylenediamine (mPDA), was initially used to analyze the kinetics of oxime ligation on aldehyde- and ketone-containing small molecules. While mPDA is only modestly more effective than aniline when used in equal concentrations (~ 2-fold), its much greater aqueous solubility relative to aniline allows it to be used at higher concentrations, resulting in significantly more efficient catalysis. In the context of protein labeling, it was first used to site-specifically label an aldehyde-functionalized protein through oxime ligation, and its kinetics were compared to reaction with aniline. Next, a protein was labeled with an aldehyde-containing substrate in crude cell lysate, captured with hydrazide-functionalized beads and then the kinetics of immobilized protein release via hydrazone-oxime exchange were analyzed. Our results show that mPDA can release and label 15 times more protein than aniline can in 3 h. Then, using the new catalyst, ciliary neurotrophic factor, a protein with therapeutic potential, was successfully labeled with a fluorophore in only 5 min. Finally, a protein containing the unnatural amino acid, *p*-acetyl phenylalanine, a ketone-containing residue, was prepared and PEGylated efficiently via oxime ligation using mPDA. This new catalyst should have a significant

impact on the field of bioconjugation, where oxime ligation and hydrazone-oxime exchange are commonly employed.

Introduction

Imine-based reactions are widely used to link complex biomolecules due to their high chemoselectivity and reversibility.[150–152,123,153–156] When there is an oxygen or nitrogen atom adjacent to a nitrogen as in the cases of hydrazines and alkoxyamines, the reaction favors imine formation;[157,141] in the absence of such a substituent, the equilibrium favors formation of free amine and free aldehyde or ketone. Among all imine-based reactions, oximes are the most stable imines[141] and as a result have found widespread use in applications such as protein labeling,[11–15] analyzing protein-protein interactions and *in vivo* cell imaging.[9,158] Although aniline is known to catalyze the oxime formation reaction, it suffers from relatively slow reaction kinetics, especially when a ketone is being used[11,18] or when transoximization is conducted.[39] In the case of ketones under physiological conditions, even in presence of aniline, it takes several hours for the reaction to be complete although faster rates can be achieved if higher reactant concentrations are employed. Moreover, in the case of hydrazone-oxime exchange, the reaction rates are orders of magnitudes slower. Therefore, discovery of an improved catalyst would significantly improve the utility and broaden the scope of this valuable bioorthogonal reaction. Here, we introduce a new catalyst, *m*-phenylenediamine (mPDA) that can accelerate oxime ligation several times faster compared with aniline-catalyzed oxime ligation. That feature, in concert with the substantially greater aqueous solubility

of *m*-phenylenediamine allows rate accelerations of approximately 15-fold to be obtained.

Experimental Section

Dansyl fluorescence assay. Reaction mixtures contained phosphate buffer (PB) (100 mM, pH 7.0), varying concentrations (10 μM to 300 μM) of aminooxy-dansyl (**1**), 0.08 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 50 μM aldehyde (citral or dodecanal) and varying concentrations of catalysts, in a final volume of 250 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the aldehyde, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 50 min.

Catalyst screening. *m*-Phenylediamine, *o*-phenylenediamine, *p*-phenylediamine, *o*-aminophenol, *m*-aminophenol, *p*-aminophenol, *o*-aminobenzoic acid and aniline were analyzed for catalytic activity in the oxime ligation reaction. Reaction mixtures contained PB (100 mM, pH 7.0), 100 μM aminooxy-dansyl (**1**), 0.08 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 50 μM aldehyde (citral or dodecanal) and 25 μM catalyst, in a final volume of 200 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the aminooxy reagent, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for 50 min.

Effect of the catalyst concentration on the k_{obs}. Reaction mixtures contained PB (100 mM, pH 7.0), 100 μM aminooxy-dansyl (**1**), 0.08 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 30 μM aldehyde (citral) and varying concentrations of catalysts (25 to 50 mM) in a final volume of 200 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by

the addition of the aldehyde, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 25 min.

Kinetic analysis of oxime ligation between 2-pentanone and aminooxy-dansyl 1. Reaction mixtures contained Tris·HCl (50 mM, pH 7.5), aminooxy-dansyl **1** (150 μM), 0.4 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 5 mM ketone (2-pentanone) and varying concentrations of catalysts, in a final volume of 200 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the aminooxy, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 4 h.

Enzymatic incorporation of 2 into GFP-CVIA (3). Enzymatic reaction mixtures (50 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), GFP-CVIA **(3,** 2.4 μM), **2** (50 μM), and PFTase (200 nM). After incubation at rt overnight, the reaction mixture was concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess **2** was removed with a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluent. The subsequent protein concentration was calculated by UV absorbance at 488 nm (ε=55,000 $M^{-1}\cdot cm^{-1}$).

Kinetic analysis of protein labeling via oxime ligation. Reaction mixtures contained PB (100 mM, pH 7.0), 10 μM GFP-aldehyde **4a**, 50 μM aminooxy-dansyl **1** and varying concentrations of *m*-phenylenediamine or 100 mM aniline, in a final volume of 250 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the catalyst, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 100 min.

Crude prenylation, immobilization and subsequent labeling and release of GFP-CVIA. A pellet of cells expressing GFP-CVIA were suspended in buffer (20 mM Tris·HCl pH 7.5, 1 mM EDTA), sonicated and clarified by centrifugation. The GFP concentration present in the crude soluble protein mixture was calculated by UV absorbance at 488 nm. Next, prenylation was performed by adding PFTase (200 nM), **2** (50 µM), Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 µM) and DTT (5.0 mM) to a solution of **3,** to achieve a final concentration of 2.0 µM in the crude mixture. After overnight incubation at rt, the reaction mixture was filtered and concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess **2** was removed through a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluting solvent. The subsequent GFP concentration in the crude mixture was calculated by UV absorbance at 488 nm. Hydrazide agarose beads (Thermo Scientific, hydrazide loading: 16 µmol/mL) (300 µL) were washed with PB (0.1 M, 3x500 µL, pH 7.0). PB (30 µL, 1 M, pH 7.0) was added to the beads followed by addition of **4a** in the crude mixture (200 µL, 70 µM). Immobilization was initiated by adding aniline (100 mM) or *m*-phenylenediamine (40 mM). After 1 h with constant agitation, the beads were washed thoroughly with PB (0.3 M, pH 7.3) and KCl (3x300 µL, 1 M) to remove non-specifically bound proteins followed by incubation with aminooxy fluorophore **5** (0.7 mM) and either *m*-phenylenediamine (700 mM) or aniline (100 mM) with constant agitation. The supernatant was then analyzed via SDS-PAGE and in-gel fluorescence analysis to compare the amount of protein released from the beads with either catalyst.

LC-MS analysis of proteins for determination of prenylation and labeling efficiency. Purified prenylated GFP (**4a**) and pure GFP-CVIA (**3**) were analyzed by LC-MS to ensure complete prenylation. Proteins were stored in Tris·HCl (50 mM, pH 7.5) prior to injection into the LC-MS instrument. Crude reaction mixtures of GFP-aldehyde **4a** and aminooxy **1** catalyzed by either aniline or *m*-phenylenediamine were analyzed by LC-MS to ensure complete ligation in both cases of the catalysts. The LC-MS method used was a gradient of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

Kinetic analysis of hydrazone-oxime exchange. The hydrazide beads containing immobilized GFP were incubated in PB (0.3 M, pH 7.0) with aminooxy-alexafluor-488 **5** (1 mM) and either aniline (100 mM) or *m*-phenylenediamine (750 mM) with constant agitation of the resulting mixtures at rt. The solutions were centrifuged at different time points (every 1 h for mPDA and every 2 h for aniline for a total of 8 h) and the amounts of released protein in the solutions were analyzed via SDS-PAGE. Gels were visualized by staining with Coomassie blue after being scanned via in-gel fluorescence imaging of alexafluor-488.

PEGylation from immobilized GFP-beads. Immobilization was performed as described above. Beads were washed thoroughly with PB (0.3 M, pH 7.3) and KCl (3x300 µL, 1 M) to remove non-specifically bound proteins. Next, the beads were incubated with aminooxy PEG **10** (5 mM) and *m*-phenylenediamine (200 mM) for 2 h with constant agitation of the solution at rt. MALDI-MS analysis of the supernatant indicated the successful PEGylation and release of the aldehyde-GFP from the hydrazide-beads.

Prenylation of CNTF-CVIA with aldehyde analog 2. Enzymatic reaction mixtures (20 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), CNTF-CVIA (2.4 μM), **2** (50 μM), and PFTase (200 nM). After incubation at rt for 90 min, the reaction mixture was concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess **2** was removed with a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluent. Purified prenylated CNTF (**9**) and pure CNTF-CVIA were analyzed by LC-MS to ensure complete prenylation. Proteins were stored in Tris·HCl (50 mM, pH 7.5) prior to injection into the LC-MS instrument.

Coupling reaction between aldehyde-labeled CNTF-CVIA (9) with alexafluor-488 (5). Alexafluor-488 (**5**) (4.2 μL of 3.2 mM solution in DMSO) was added to 42 μL of **9** (stock solution of 60 μM in Tris·HCl (50 mM, pH 7.5)). PB (2 M, pH 7, 2.5 μL) was added and the reaction was initiated by adding 50 mM *m*-phenylenediamine (stock solution of 1.5 M in 0.3 M PB, pH 7.0) and was allowed to proceed for 30 min at rt. LS-MS analysis of the sample showed only oxime ligated protein and no free aldehyde was detected indicating a complete reaction in both prenylation and oxime ligation reactions.

Rate analysis of the coupling reaction between aldehyde-CNTF 9 and aminooxy 5. Alexafluor-488 (**5**) (8 μL of 3.1 mM solution in DMSO) was added to 6 μL of PB (1 M, pH 7) and 2 μL of *m*-phenylenediamine (stock solution of 1.5 M in 0.3 M PB, pH 7.0) or 2 μL water as the control reaction. Reactions were performed at rt and were initiated by adding 20 μL of **9** (stock solution of 60 μM in Tris·HCl (50 mM, pH 7.5)). To monitor the reactions, 15 μL aliquots were withdrawn at 5 min intervals, added to 15 μL of

loading buffer, flash frozen by liquid nitrogen and subjected to SDS-PAGE analysis. Samples were heated at 98 °C for 4 min prior to gel analysis.

PEGylation of DHFR2 M174pAcF 11 protein with aminooxy-PEG 12 using mPDA. $DHFR^2$ fusion protein with the unnatural amino acid *p*-acetyl phenylalanine ($DHFR^2$ M174pAcF) (7 μM) was reacted with aminooxy-PEG **12** (3 kDa) (5 mM) in PB (0.1 M, pH 7) in the presence of either 100 mM aniline, 500 mM mPDA or no catalyst at rt. The amounts of PEGylated protein in the solutions were analyzed at different time points via SDS-PAGE. Gels were visualized by staining with Coomassie blue. Densitometry analyses on the stained gels were performed using the program ImageJ v1.46.

General procedure for MALDI analysis of protein samples. The sample was adsorbed onto a zip-tip (C_4 column) via repeated cycles of aspiration and ejection (5-10 cycles of 10 μL each) using a pipettor. Next, in order to remove excess buffer and reagents, the zip-tip was washed 5x10 μL with solvent A (H_2O containing 0.1% TFA; v/v) and the proteins eluted with 2 μL of a mixture of A and B (25:75) (solvent B: CH_3CN containing 0.1% TFA; v/v). Next 0.7 μL of the eluted material was added to a MALDI plate and 0.7 μL of matrix was added on top of the sample plate and both were mixed thoroughly to form crystals. A saturated solution of sinapinic acid (3,5-dimeth-oxy-4-hydroxy-cinnamic acid) was used as the matrix.

Results and Discussion

Development of an assay to analyse kinetic of the oxime ligation reaction.

In order to study catalysis of oxime ligation, a continuous fluorescence assay was developed that enabled easy monitoring of the reaction. The dansyl chromophore manifests environmentally sensitive fluorescence that increases in the presence of hydrophobic groups.[159] Thus, it was postulated that reaction of an aminooxy dansyl compound with a hydrophobic aldehyde or ketone would result in an increase in dansyl fluorescence that could easily be used to monitor the kinetics of oxime ligation. Therefore, an aminooxy-derivatized dansyl-containing compound, **1,** was designed and synthesized in three steps (Figure 4.1A). Incubation of **1** in the presence of citral, dodecanal or 2-pentanone resulted in a time dependent increase in fluorescence indicative of oxime formation. The formation of these products was confirmed by HPLC and MS analysis.

Screening for a new catalyst suitable for oxime ligation reaction.

After validating the continuous fluorescence assay, a number of different compounds were examined to identify new putative oxime ligation reaction catalysts. Two key features were considered in screening for such catalysts. First, the basicity of the intermediate Schiff base imine formed[142] should not be significantly lower than the original amine (catalyst) basicity. Second, the catalyst should have high water solubility. Aromatic amines that are conjugated, such as aniline, meet only the first requirement in that their Schiff bases have basicities close to their free amines.[160] Based on the idea that

the presence of electron donating groups on an aromatic ring would render the corresponding Schiff base more basic, it was reasoned that such compounds would be

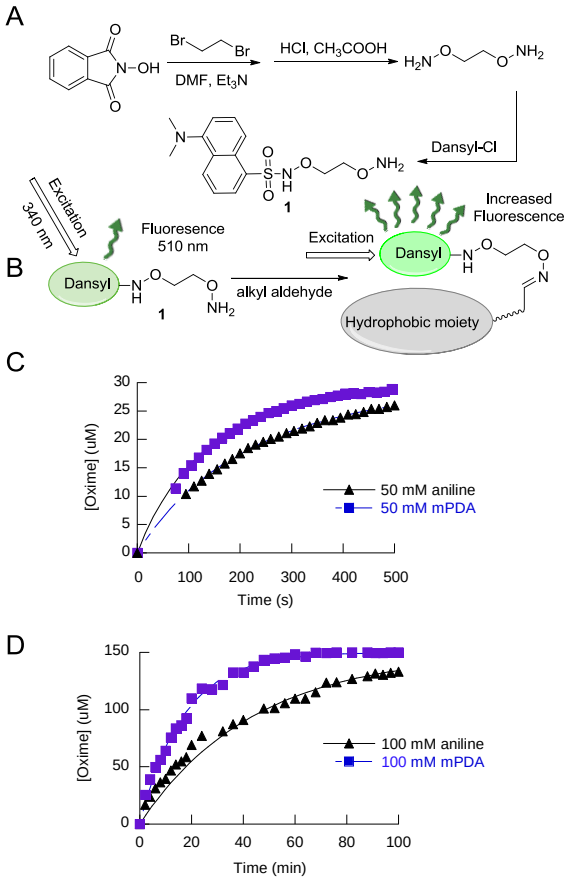


Figure 4.1. A) Synthesis of dansyl-aminooxy **1**. B) Schematic representation of dansyl-fluorescence assay. When the dansyl moiety is in close proximity to a hydrophobic group, its fluorescence increases. "Hydrophobic moiety" represents the alkyl chain of the aldehyde- or ketone-containing reactants. C) Analysis of the reaction of **1** (100 µM) with citral (30 µM), a conjugated hydrophobic aldehyde, in presence of either 50 mM mPDA (squares) or 50 mM aniline (triangles) with k_{obs} of 78.2 $s^{-1}M^{-1}$and 48.9 $s^{-1}M^{-1}$ respectively. D) Analysis of the reaction of **1** (150 µM) with 2-pentanone (5 mM), a non-conjugated ketone, in presence of either 100 mM mPDA (squares) or 100 mM aniline (triangles) with k_{obs} of 0.20 $s^{-1}M^{-1}$and 0.082 $s^{-1}M^{-1}$ respectively.

more efficient reaction catalysts. Thus, *o*-, *m*- and *p*-aminophenols, *o*-aminobenzoic acid and *o*-, *m*- and *p*-phenylenediamines were examined. The latter three compounds have the advantage of having two amino groups that theoretically increases the probability of formation of an intermediate Schiff base that could result in higher catalytic efficiency. After using the fluorescence assay described above to screen all seven proposed catalysts, and fitting the data to the second order kinetic model previously employed by Dawson and coworkers,[142] a range of results were obtained and are summarized in Table 4.1 and Figure 4.S1. Given the variations in rate observed with the different catalysts, it is likely that a complex interplay of factors including inductive, steric and resonance effects together with hydrogen bonding (in the case of *ortho* substituents) participate in controlling their reactivity. *o*-Aminophenol was found to have a solubility too low to be useful. *o*-Phenylenediamine, *m*-aminophenol, and *o*-aminobenzoate were found to be equal to or less efficient than aniline while *p*-aminophenol, *m*-phenylenediamine and *p*-phenylenediamine were found to be more efficient compared with aniline. Of those latter three, *m*-phenylenediamine (mPDA) was the most efficient catalyst (1.7-fold compared to aniline) and had high water solubility. Hence mPDA was selected for further study.

Kinetic analysis of oxime ligation reactions using aniline and *m*-phenylenediamine (mPDA).

Next, the reactions between several aldehyde- and ketone-containing compounds and alkoxyamine **1** were studied using aniline and mPDA as catalysts. Two hydrophobic aldehydes compatible with the fluorescence-based assay, citral and dodecanal (conjugated and nonconjugated aldehydes, respectively) were selected for initial

experimentation. Analyses were performed using 50 mM catalyst (aniline or mPDA), 100 µM **1** and 30 µM aldehyde in phosphate buffer at pH 7.3. Interestingly, analysis of those reactions showed that both aldehydes were labeled with dansyl-aminooxy **1** within a few minutes. An equal concentration of mPDA was almost twice as efficient (Figure 4.1C, blue squares) as aniline (black triangles) consistent with what was observed in the initial screening; when no catalyst was employed, oxime formation required more than 80 min to achieve comparable levels of conversion under the same conditions (Figure 4.S2-A, green circles). Significantly, citral, a conjugated aldehyde, reacted almost two times slower than dodecanal did (Figure 4.S2-B). Given the importance of ketone-containing compounds as targets for bioorthogonal reactions, 2-pentanone was also studied as an example and was found to react at a rate at least two orders of magnitude slower than either of the two aldehydes (k_{obs} for oxime ligation of citral and 2-pentanone with aminooxy-dansyl **1** were 48.6 $s^{-1}M^{-1}$ and 0.082 $s^{-1}M^{-1}$ using 50 mM and 100 mM aniline as a catalyst, respectively; Table 4.1 and Table 4.3). This highlights the role of carbonyl group reactivity in oxime formation kinetics. Aldehydes react more rapidly than ketones and non-conjugated aldehydes are more reactive than conjugated ones. Dawson and coworkers originally established that the kinetics of oxime ligation reactions catalyzed by aniline fit a second order model that is first order in both aldehyde and alkoxyamine.[142] They also showed that the apparent second order rate constant for this process varied with aniline concentration. Wen-jun et al. extended those observations by demonstrating that the apparent second order rate constant varied linearly with aniline.[161] Accordingly, we analyzed the rate data for the citral model reaction described here and confirmed that both

the aniline- and mPDA-catalyzed reactions vary linearly with catalyst concentration (Figure 4.S4 and 4.S5; Table 4.2). A linear dependence on reaction rate with catalyst concentration was also observed in the model reaction involving the ketone, 2-pentanone (Table 4.3 and Figure 4.S3). Overall, the results of these model studies indicate that the increase in reaction rate using mPDA is approximately 2-fold higher than that obtained with aniline when the catalysts are employed at equal concentrations. However, the fact that the rate of oxime formation is first order in catalyst, coupled with the much greater solubility of mPDA suggested that it should be possible to obtain much greater rate acceleration with mPDA by employing it at concentrations substantially above the solubility limit of aniline (~ 100 mM). That feature is apparent from the data shown in Table 4.3 for 2-pentanone and was investigated in more detail using aldehyde- and ketone-containing proteins as described below.

Kinetic analysis of oxime ligation on proteins.

To examine the kinetics of oxime ligation on proteins, the enzyme protein farnesyltransferase (PFTase) was employed to introduce aldehyde functionality into proteins. In previous work, a number of groups have exploited the high specificity of PFTase to site-specifically modify proteins. PFTase catalyzes the transfer of an isoprenoid group from farnesyl pyrophosphate (FPP, Figure 4.2A) to a cysteineyl sulfur atom present in a tetrapeptide sequence (denoted as a CaaX-box) positioned at the C-terminus of a protein. Importantly, CaaX-box sequences such as CVIA can be appended to the C-termini of many proteins rendering them efficient substrates for PFTase. Since PFTase can tolerate many simple modifications of the isoprenoid substrate, it can be used

to introduce a variety of different functional groups into proteins; chemoselective reaction with the resulting functionalized protein can then be used for a wide range of applications. We previously showed that aldehyde-containing FPP analogues could be

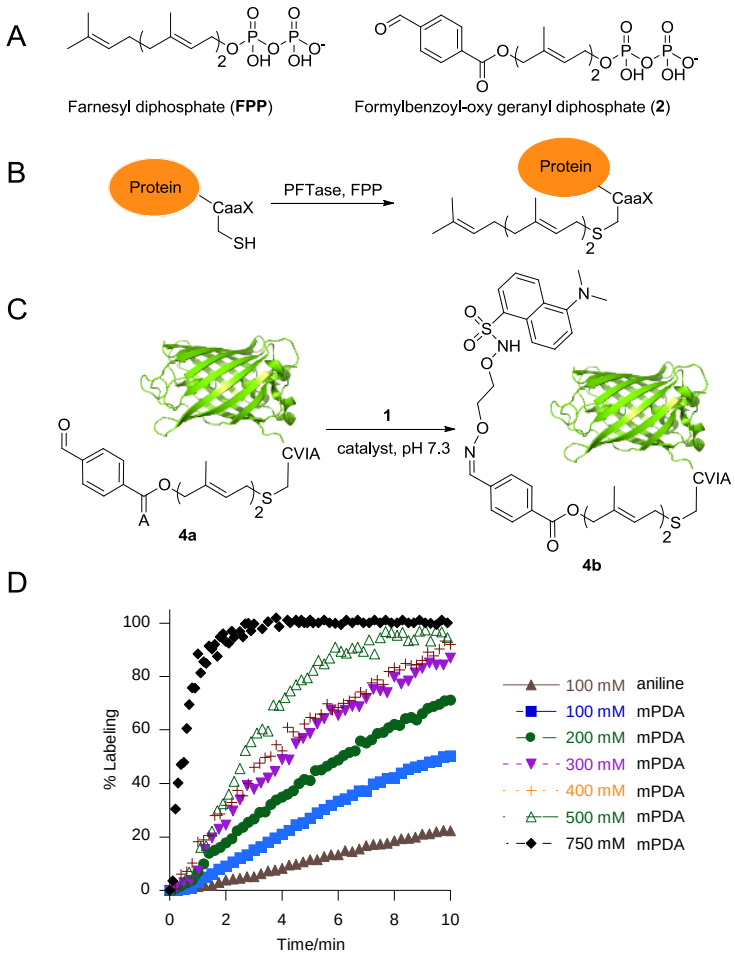


Figure 4.2. Kinetic analysis of oxime ligation with an aldehyde-functionalized protein. A) Structures of FPP and compound **2**. B) Schematic representation of a prenylation reaction with a protein containing a CaaX box at its C-terminus. C) Schematic representation of oxime ligation with an aldehyde-functionalized protein **4a** using either aniline or mPDA as the catalyst. D) Kinetic analysis of oxime ligation between **4a** (10 μM) and **1** (50 μM) using either aniline or different concentrations of mPDA. Note: at high mPDA/aminooxy ratios, Schiff base formation is competitive with oxime ligation. Hence, the % labeling reported here for oxime formation is relative to the value at

equilibrium and not complete conversion (*vide infra*). The solubility limit for aniline is approximately 100 mM under these conditions.
successfully incorporated into the proteins using this strategy and used such aldehyde-modified proteins here.

GFP-CVIA (**3**) was employed as a model protein to perform oxime kinetic analysis (Figure 4.2C). The protein was tagged with analogue **2**[40] using PFTase to yield aldehyde-labeled protein **4a** that was used in subsequent ligation reactions. Dansyl-aminooxy **1** was added to a solution of 10 μM aldehyde-functionalized protein **4a** and the reactions were initiated by the addition of either aniline or mPDA catalyst. As was observed with the simple aldehyde model compounds described above, oxime formation between **1** and **4a** resulted in a significant increase in dansyl group fluorescence that was then used to measure the reaction rate; no fluorescence change was observed when unprenylated GFP-CVIA (**3**) was treated with **1** (Figure 4.S7). Additionally, analysis of the initial protein **3**, aldehyde-functionalized protein **4a**, and oxime-labeled protein **4b** via LC-MS gave species at 27,335 Da, 27,619 Da, 27,926 Da, confirming that prenylation and the following oxime labeling were proceeding as expected (Figure 4.S6). Interestingly, examiniation of the reaction rates using **4a** via the fluorescence assay showed that both catalysts were somewhat less effective with the protein aldehyde than they were with the simple model aldehydes (Figure 4.S7). That observation suggests that the reactivity of the aldehyde can be modulated by the protein to which it is attached and is consistent with previous observations reported by Dawson and Bertozzi where varying extents of acceleration by aniline were reported.[153,158] Since mPDA, in contrast to aniline, has a high solubility in water at pH 7 (approximately 100 mM for aniline compared to >2 M for

mPDA), we decided to examine the effect of employing higher concentrations of the new catalyst on the oxime ligation reaction using a protein-aldehyde substrate. Consistent with the first-order kinetics observed in the model reactions discussed above, a plot of rate versus mPDA concentration for oxime-protein ligation was also linear (Figure 4.S8). Thus, the ability to employ higher concentrations of mPDA compared with aniline due to its greater solubility resulted in significant rate acceleration; for example inclusion of mPDA at 750 mM resulted in essentially complete labeling of the aldehyde-functionalized protein **4a** in about 90 s (Figure 4.2D), which is an impressive result compared with the minimal reaction (<7%) obtained with 100 mM aniline during that same time period. Kinetic analysis revealed that when aniline and mPDA were used at the same concentrations (100 mM), mPDA catalyzed the reaction ~2.5 times more efficiently than aniline (compare k_{obs} of 10.3 $M^{-1}\cdot s^{-1}$ for aniline versus 27.0 $M^{-1}\cdot s^{-1}$ for mPDA, Table 4.S1). More importantly the rate was greater than 10 times higher when mPDA was 500 mM and approximately 15 times higher when mPDA was 750 mM (Figure 4.S8 and Table 4.S1).

Capture and release strategy to purify and label proteins from crude cell extract using mPDA. For many protein labeling applications, the target polypeptide is not always pure or present in high abundance; therefore, specific modification strategies that function in a crude mixture are necessary. In previous work, we reported on a method for site-specifically functionalizing a protein in a crude cell mixture, and labeling and purifying the modified protein using aniline without any additional purification.[40] Although that process was successful, it required long reaction times to achieve

significant levels of conversion. Accordingly, we next re-evaluated our aforementioned method for protein labeling with the new, more effective, mPDA catalyst discovered

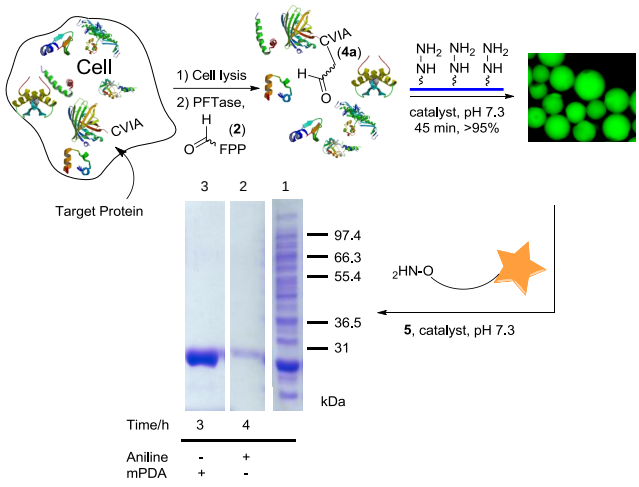


Figure 4.3. Purification and selective labeling of a protein using mPDA. Target protein was site-specifically tagged by aldehyde-FPP analog via PFTase followed by capture of the aldehyde-functionalized protein in the crude cell lysate by hydrazide functionalized surfaces. Prenylation in the crude was confirmed by LC-MS analysis on the crude extract after prenylation. The immobilized protein was then released into the solution and fluorescently labeled by addition of aminooxy-fluorophore **5** (0.7 mM), using either mPDA (700 mM) or aniline (100 mM, its approximate solubility limit) as the catalyst. SDS-PAGE analysis (visualized by staining with Coomassie blue): Right lane (1): crude cell lysate, middle lane (2): released protein (after immobilization) obtained using aniline for 4 h; Left lane (3): released protein after immobilization obtained using mPDA for 3 h.

here. Thus, *E. coli* cells expressing GFP-CVIA **3** were lysed and enzymatically prenylated using PFTase and substrate **2**. LC-ESI/MS analysis of the reaction sample confirmed incorporation of the aldehyde functionality into the GFP-CVIA **3** in the crude cell lysate. The reaction mixture was then concentrated and excess **2** was removed via size exclusion column chromatography. Aldehyde-GFP **4a** was then selectively immobilized from the crude cell lysate onto hydrazide-functionalized beads using either

100 mM aniline or 40 mM mPDA as the catalyst. Immobilization was complete in less than 45 min in both cases at which point the beads became highly fluorescent. Next, the

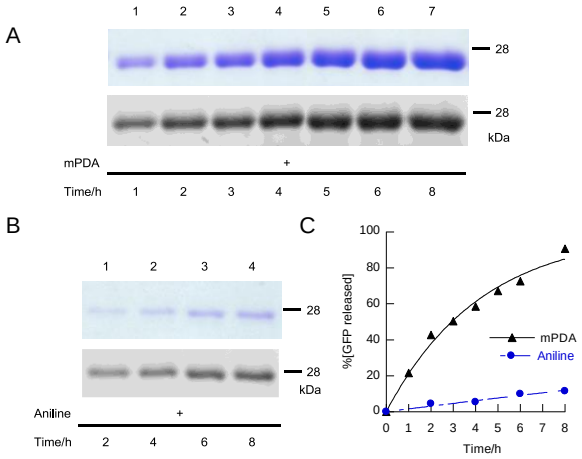


Figure 4.4. Kinetic analysis of the release of the hydrazone immobilized protein into fluorescently labeled oxime protein via hydrazone-oxime exchange reaction. Immobilized protein was incubated with aminooxy fluorophore **5** (1 mM) and catalyst, followed by analysis of the amount of released protein in the solution via SDS-PAGE. In (A) and (B), the upper bands were visualized by staining with Coomassie blue while the lower bands were detected via in-gel fluorescence imaging of Alexafluor-488. A) Time course of protein release in 1-8 h using mPDA (750 mM) as the catalyst. B) Time course of protein release using aniline (100 mM, its approximate solubility limit). C) Comparison of protein release rates using aniline or mPDA obtained by densitometric analysis of the SDS-PAGE results shown in (A) and (B) using the Coomassie blue stained gels.

beads were thoroughly washed to remove any non-specifically bound proteins and then treated with aminooxy-alexafluor-488 **5** in the presence of 100 mM aniline or 700 mM mPDA and the amount of released and fluorescently labeled protein was measured. SDS-PAGE analysis of the supernatant solution showed a single band attributable to the released, labeled GFP **4d** (Figure 4.3). Qualitatively, from that data, it is clear that the

efficiency of protein release using mPDA is substantially greater than with aniline. A more thorough kinetic analysis revealed that mPDA, under these reaction conditions, can release immobilized protein from the beads with an initial velocity approximately 15 times faster than aniline (compare k_{obs} of 0.0159 h^{-1} for 100 mM aniline with 0.237 h^{-1} for 750 mM mPDA, Table 4.S2, Figure 4.4). The in-gel fluorescence analysis further confirmed that the released protein was actually labeled with fluorophore **5** with either catalyst (Figure 4.4). Overall, this important result indicates that a protein can be enzymatically modified in a crude mixture, selectively immobilized onto hydrazide beads in the presence of many other proteins and then successfully labeled and released back into the solution in high yield using an aminooxy-fluorophore in less than 8 h. It is worth noting that when the mPDA/aminooxy reagent ratio is very high (>250), Schiff base formation (as an end product) is competitive with oxime ligation. Thus, we recommend that mPDA/aminooxy reagent ratios of less than 250 be used so that competitive Schiff base formation becomes negligible.

Capture and release strategy to purify and PEGylate proteins from crude cell extract using the catalyst. The attachment of polyethylene glycol (PEG) chains to proteins is the most widely employed method for increasing the half-life of protein-based therapeutic agents in blood.[143,145] Site-specific methods for protein PEGylation can minimize deleterious effects associated with nonselective PEGylation.[162] In previous work, we described a method for site-specific protein PEGylation based on the capture and release strategy outlined above using aniline as a catalyst.[40] In that work, PEGylation was achieved by releasing the captured protein (obtained by enzymatic aldehyde

incorporation in crude extract followed by immobilization via hydrazide formation) using an aminooxy-functionalized PEG polymer. Given the significant rate enhancement

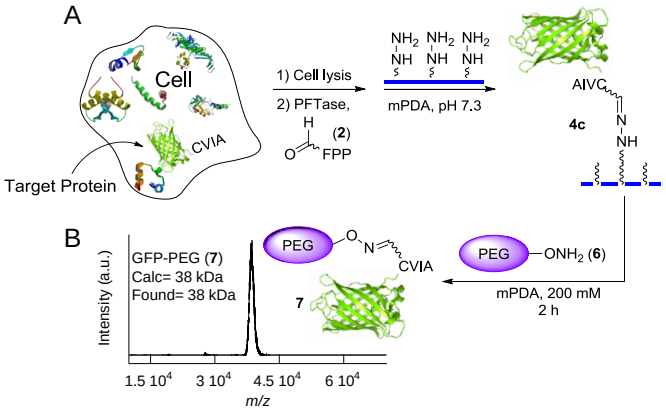


Figure 4.5. A) Use of PFTase-catalyzed protein modification for site-specific PEGylation from crude cell lysate using the mPDA. The target protein was prenylated, captured using hydrazide beads and then simultaneously released into the solution and PEGylated by the addition of aminooxy-PEG (**6**, 5 mM using 200 mM mPDA as a catalyst. B) MALDI analysis of the released material in less than 2 h confirmed the formation and release of the highly pure PEGylated GFP (**7**) into the solution.

observed with mPDA, we elected to evaluate the utility of this new catalyst for rapid protein PEGylation. Thus, **3** was prenylated with **2** using PFTase in crude *E. coli* extract followed by capture using hydrazide-functionalized agarose. After washing the material to remove nonspecifically bound proteins, the desired PEGylated protein (**7**) was eluted via treatment with 5 mM aminooxy-PEG **6** in the presence of 200 mM mPDA (Figure 4.5). Analysis of that material by MALDI-MS (Figure 4.5) showed a product with a mass of 38 kDa (and nothing else) consistent with the formation of highly pure PEGylated protein **7** from crude lysate in less than 2 h; the broad peak observed is consistent with the attachment of a polydisperse polymer to a monodisperse protein. It is also important to note that no species resulting from the Schiff base formation of mPDA or addition of

multiple PEG chains were observed, consistent with the selective nature of the chemistry employed here.

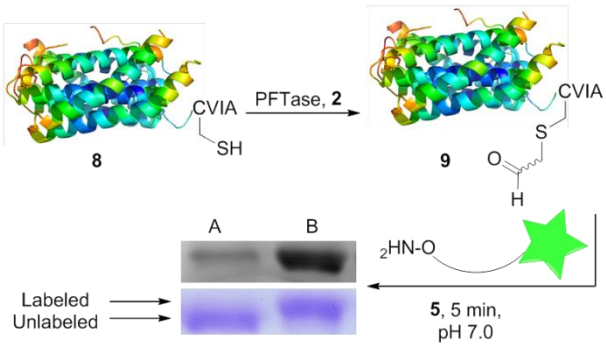


Figure 4.6. Prenylation of CNTF-CVIA (**8**) with aldehyde analog **2** followed by fluorophore labeling of the prenylated protein (50 μM) via oxime ligation with **5** (250 μM) for 5 min. A and B are SDS-PAGE analysis of the labeled protein. A) No catalyst was used for labeling. B) mPDA was used as the catalyst (80 mM) for labeling. Upper bands are detected via in-gel fluorescence analysis and lower bands are visualized by staining with Coomassie blue. The fluorescently labeled protein has a higher mass and hence shows a decrease in its electrophoretic mobility compared to the precursor prenylated protein.

Rapid labeling of CNTF, a protein of biomedical importance, using mPDA. Having established the utility of this method for rapid C-terminal site-specific modification with a model protein, GFP, we decided to illustrate its use by modifying a protein of biomedical importance. Ciliary neurotrophic factor (CNTF) is a member of a class of proteins that stimulate neurite outgrowth and promote neuron survival during inflammatory events.[163] Recently, CNTF has been studied extensively as a possible therapeutic agent for slowing retinal degeneration.[164,165] Thus, we elected to investigate the preparation of a fluorescent form of CNTF using the new catalyst reported here. To accomplish this, purified CNTF−CVIA (**8**), a form of CNTF engineered to contain a C-

terminal CAAX box, was prenylated with analog **2** under conditions established above for GFP followed by labeling using aminooxy fluorophore **5** and mPDA as the catalyst. LC-MS analysis (Figure 4.S9) confirmed successful prenylation and subsequent labeling of CNTF; as expected, both the enzymatic prenylation and the subsequent oxime ligation proceed with essentially complete conversion. Next, we analyzed how fast the protein is indeed labeled when the new catalyst is used. The aldehyde-labeled protein **9** (50 μM) was incubated with alkoxyamine **5** (600 μM), with or without mPDA as the catalyst (80 mM) and the reaction was monitored via SDS-PAGE as a function of time. In-gel fluorescence analysis of the resulting gel showed a substantially higher labeling efficiency when mPDA was used as the catalyst relative to when no catalyst was used (Figure 4.6). Inspection of the Coomassie blue stained gel revealed that the labeled protein manifests a decrease in electrophoretic mobility compared to the prenylated protein due to their mass difference allowing the two forms to be distinguished. Accordingly, Coomassie blue visualization of the gel showed that while almost all of the protein was labeled when mPDA was used, the amount of labeling that was obtained in the absence of mPDA was very small when the reaction was performed within the same time period (Figure 4.6).

PEGylation of a protein containing a ketone using mPDA.

In addition to aldehydes, ketone-containing proteins have also been used for a number of bioconjugation applications including protein PEGylation and spin-labeling.[166] The incorporation of the non-natural amino acid *p*-acetyl phenylalanine into polypeptides via suppressor t-RNA technology is a versatile approach for site-selective introduction of

ketones into proteins.[25,134] However, given the attenuated reactivity of ketones, oxime formation is typically carried out under conditions more extreme than those used for

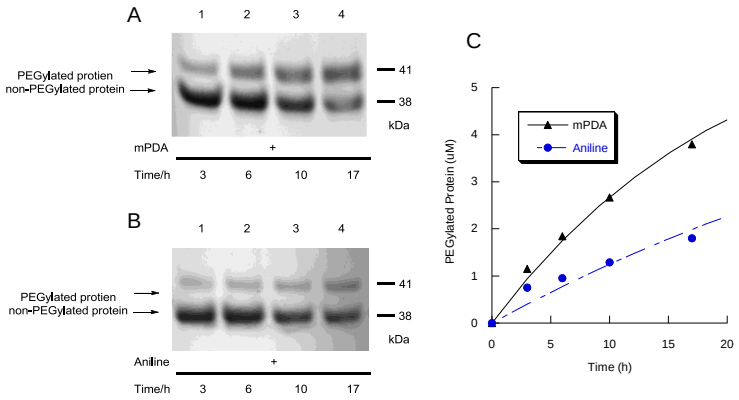


Figure 4.7. Kinetic analysis of PEGylation of the protein DHFR[2] M174pAcF which contains the unnatural amino acid *p*-acetyl phenylalanine with aminooxy-PEG (3 kDa). Protein (7 μM) was incubated with aminooxy-PEG (5 mM) with either 100 mM aniline or 500 mM catalyst. Next the samples were analyzed via SDS-PAGE and the protein was visualized by staining with Coomassie blue. A) Time course of protein PEGylation in 1-17 h using mPDA (500 mM) as the catalyst. B) Time course of protein PEGylation using aniline (100 mM, its approximate solubility limit) as the catalyst. C) Comparison of protein PEGylation rates using aniline or mPDA obtained by densitometric analysis of the SDS-PAGE results shown in (A) and (B) using the Coomassie blue stained gels.

comparable reactions with aldehydes. For example, a recent report from Schultz and coworkers employed 100 μM of a ketone-containing antibody and 5 mM alkoxyamine at pH 4.5, 37 °C to achieve near completion in 24 h.[167] To probe the utility of mPDA for catalyzing oxime formation with a ketone-containing protein, a *p*-acetyl phenylalanine-containing protein, (DHFR[2] M174pAcF, **11**) was produced in *E. coli* and purified. Aminooxy-PEG **12** (3 kDa, 5 mM) was reacted with DHFR[2] M174pAcF (**11**, 7 μM) in PB (0.1 M, pH 7) in the presence of either 100 mM aniline, 500 mM mPDA or no catalyst. The extent of PEGylation was analyzed at different times via SDS-PAGE and

the protein bands were visualized by staining with Coomassie blue. Densitometric analysis of the gels revealed that while aniline did not have a significant effect on the reaction rate (Figure 4.S10), approximately 2.5 fold increase in the rate was observed when mPDA was used as the catalyst (Figure 4.7). It should be noted that while the protein concentration in this experiment (7.0 μM) was comparable to that used in the work described above with aldehyde-containing proteins (10 μM), it was necessary to increase the concentration of the aminooxy-PEG reagent to compensate for the decreased reactivity of the ketone (compare 0.1-1.0 mM for the aldehyde with 5 mM for the ketone). The presence of elevated concentrations of the aminooxy compound increases the rate for uncatalyzed oxime formation thereby reducing the net effect of the catalyst. Nevertheless, using mPDA, it was still possible to obtain substantial oxime formation under conditions significantly milder (7 μM protein-ketone, 5 mM aminooxy-PEG, pH 7.0, rt) than those reported by other groups.[166–168]

Effect of mPDA on protein structure and function. Since the mPDA catalyst described here is typically employed at elevated concentrations (relative to aniline), we decided to study its effect on protein structure and enzymatic function using two different assays. In the first assay, GFP was treated with different concentrations of mPDA followed by gel filtration chromatography to remove the catalyst. Circular dichroism spectroscopy of the resulting samples showed no significant differences suggesting that exposure to mPDA does not cause substantial denaturation or irreversible protein unfolding (Figure 4.S12). In the second assay, the effect of mPDA on enzymatic activity was examined. PFTase was treated with various concentrations of mPDA followed by dilution (to reduce the

mPDA concentration by a factor of 20) and activity measurement using the aforementioned spectrofluorometric assay. Analysis of that data showed no significant loss of enzymatic activity following treatment with mPDA (Figure 4.S8). Overall, these results suggest that mPDA can be used to catalyze oxime-ligation reactions with proteins without deleterious effects on protein structure or enzymatic activity.

Conclusion. In conclusion, an analysis of the oxime ligation reaction for bioconjugation purposes was performed and resulted in the discovery of several new catalysts for oxime formation and hydrazone-oxime exchange with, *m*-phenylenediamine (mPDA), being the most efficient. mPDA is a highly efficient catalyst that accelerates oxime formation by orders of magnitude under physiological conditions; compared with aniline, it can catalyze the reaction approximately two-fold faster when employed at equal concentrations. Importantly, mPDA has the advantage of being significantly more soluble than aniline. That feature, in concert with greater catalytic efficiency, allows oxime ligation and hydrazone-oxime exchange to be performed up to 15-fold faster with mPDA compared with aniline; this result is reminiscent of observations made in mechanistic studies of native chemical ligation where replacement of thiophenol with 4-carboxymethyl thiophenol led to significant rate enhancement in transthioesterification due to the much greater solubility of the latter compound.[169] The acceleration obtained with mPDA is particularly useful for bioconjugation where high concentrations of proteins cannot typically be obtained to increase the reaction rate. To showcase the utility of this new catalyst, we demonstrated that a protein can be enzymatically labeled with an aldehyde moiety in crude cellular extract, selectively immobilized onto hydrazide beads,

and released back into the solution as a fluorescently labeled or PEGylated protein via mPDA catalyzed hydrazone-oxime exchange in a few hours. The amount of protein released in 3 h from the beads was up to 15 times higher in the case of mPDA compared to aniline as the catalyst. While the decreased reaction times achievable with mPDA could be useful in many contexts, this catalyst may be particularly useful for the preparation of materials containing short half-life radionuclides employed for various biomedical imaging applications.[170–172] Overall, this new catalyst should be useful for a wide range of bioconjugation applications that require rapid reactions under mild, biocompatible reaction conditions.

Funding Sources. This work was supported by the National Institutes of Health (GM058842 and GM084152), University of Minnesota Endowment Funding and the Minnesota Supercomputer Institute. We also thank Dr. Peter Schultz's group at Scripps Research Institute for providing the plasmid pEVOL_pAcF.

Table 4.1. Kinetic analysis of oxime ligation reactions between the aldehyde, citral, and **1** with different catalysts.

Catalyst[a]	k_{obs} [b] (s^{-1} M^{-1})	k_{obs} /$k_{obs(\text{aniline})}$	$t_{1/2}$ (min)	Approximate Solubility[c]
Aniline	24.4±0.5	-	5.54	100 mM
o-Aminophenol	-	-	-	Insoluble
m-Aminophenol	23.9±0.5	0.98	5.65	100 mM
p-Aminophenol	33.5±0.9	1.37	4.04	50 mM
o-Aminobenzoic acid	24.8±0.8	1.01	5.45	very soluble (>2 M)
o-Phenylene-diamine	18.1±0.3	0.74	7.47	very soluble (>1 M)
m-Phenylene-diamine	41.5±1.2	1.70	3.26	very soluble (>2 M)
p-Phenylene-diamine	29.2±0.7	1.20	4.63	very soluble (>1 M)

[a]Reactions were performed using citral (50 μM), aminooxy-dansyl **1** (100 μM) and catalyst (25 mM).
[b]The k_{obs} values were obtained by fitting the experimental data to Equation S3 using Kaleidagraph v4.1.3. The values are provided ± the standard error obtained from the curve fit.
[c]Solubility in phosphate buffer (0.1 M), pH 7.0, at rt.

Table 4.2. Kinetic analysis of oxime ligation reaction, between the aldehyde, citral, and **1**, with varying concentrations of catalysts.

Aniline[a]	k_{obs} [b] ($s^{-1}\,M^{-1}$)	$k_{obs}/[cat.]$ ($ms^{-1}\,M^{-2}$)	$t_{1/2}$ (min)
25 mM	24.3±0.2	0.97	5.20
35 mM	33.5±0.2	0.96	3.77
42 mM	36.9±0.3	0.88	3.42
50 mM	48.6±0.2	0.97	2.60

mPDA[a]	k_{obs} [b] ($s^{-1}\,M^{-1}$)	$k_{obs}/[cat.]$ ($ms^{-1}\,M^{-2}$)	$t_{1/2}$ (min)	$k_{obs(mPDA)}/k_{obs(aniline)}$
25 mM	38.7±0.1	1.55	3.26	1.59
35 mM	58.3±0.3	1.66	2.17	1.74
42 mM	65.6±0.6	1.56	1.92	1.78
50 mM	78.2±0.7	1.56	1.62	1.61

[a]Reactions were performed using citral (30 μM), aminooxy-dansyl **1** (100 μM) and varying concentrations of catalysts (25 to 50 mM).
[b]The k_{obs} values were obtained by fitting the experimental data to Equation S3 using Kaleidagraph v4.1.3. The values are provided ± the standard error obtained from the curve fit.

Table 4.3. Kinetic analysis of oxime ligation reaction between the ketone, 2-pentanone, and **1** at different catalyst concentrations.

[Catalyst][a] (mM)	k_{obs} [b] ($s^{-1}\,M^{-1}$)	$k_{obs(mPDA)}/k_{obs(aniline)}$	$t_{1/2}$ (min)
100 aniline	0.082±0.0007	-	28.5
100 mPDA	0.20±0.0016	2.41	11.2
300 mPDA	0.36±0.0042	4.45	6.4
500 mPDA	0.73±0.0071	8.89	3.2
700 mPDA	1.17±0.013	14.3	2.0
900 mPDA	1.69±0.022	20.7	1.4

[a]Reactions were performed using 2-pentanone (5.0 mM), aminooxy-dansyl **1** (150 μM) and the catalysts at the concentrations indicated.
[b]The k_{obs} values were obtained by fitting the experimental data to Equation S3 using Kaleidagraph v4.1.3. The values are provided ± the standard error obtained from the curve fit.

Abbreviations. mPDA, *m*-phenylenediamine; PB, phosphate buffer; DHFR, dihydrofolate reductase; rt, room temperature; PFTase, protein farnesyl transferase; FPP, farnesyl pyrophosphate; PEG, poly ethylene glycol; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; pAcF, *p*-acetylphenylalanine; CNTF, ciliary neurotrophic factor; LC-MS, liquid chromatography-mass spectrometry; MALDI, matrix-assisted laser desorption ionization.

Supporting Information

General. All synthetic reactions were carried out at rt and stirred magnetically unless otherwise noted. TLC was performed on precoated (250 mm) silica gel 60 F-254 plates (Merck). Plates were visualized by staining with $KMnO_4$ or with a hand-held UV lamp. Flash chromatography was performed using a Biotage® instrument. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. ^{1}H NMR spectra were obtained at 500 MHz; ^{13}C NMR spectra were obtained at 125 MHz. All NMR spectra were acquired on Varian instruments at 25 °C. Chemical shifts are reported in ppm and *J* values are in Hz. Fluorescence assay data were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer. MS spectra for synthetic reactions were obtained on a Bruker BioTOF II instrument. Yeast PFTase was prepared as previously described.[173] Protein LC/MS analyses were performed using a Waters Synapt G2 Quadropole TOF mass spectrometer instrument. MALDI-MS analyses were performed with a Bruker MALDI TOF spectrometer instrument. Aminooxy alexafluor-488 (**5**) was from AnaSpec; hydrazide agarose beads were from Thermo Scientific. PEG aminooxy

10,000 MW was from NOF America Corporation. All solvents were of HPLC grade. All other reagents were from Sigma Aldrich.

Abbreviations.

mPDA, *m*-Phenylenediamine

FPP, Farnesyl diphosphate

DTT, Dithiothreitol

ESI-MS, Electrospray ionization mass spectrometry

RP-HPLC, Reversed-phase high-pressure liquid chromatography

PB, Phosphate buffer

PEG, Polyethylene glycol

DMAP, 4-Dimethylaminopyridine

PFTase, Protein farnesyl transferase

DMF, Dimethylformamide

DBU, 1,8-Diazabicycloundec-7-ene

TEA, Triethylamine

Tris, Tris(hydroxymethyl)aminomethane

EDTA, Ethylenediaminetetraacetic acid

GFP, Green fluorescent protein

CNTF, Ciliary neurotrophic factor

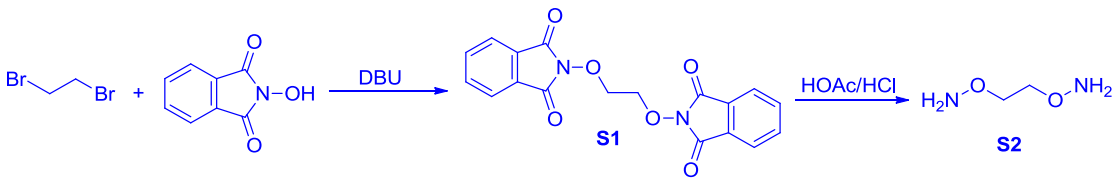
CD, Circular dichroism

DHFR, Dihydrofolate reductase

pAcF, *p*-Acetylphenylalanine

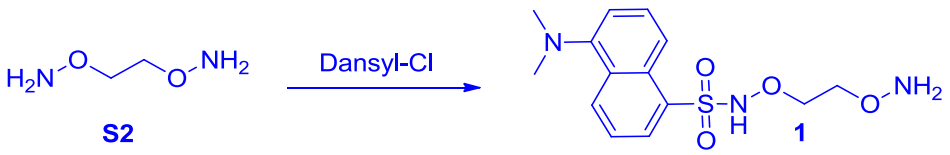
Synthesis of compound 1.

O,O'-(ethane-1,2-diyl)bis(hydroxylamine) S2.



Compound **S2** was synthesized according to a modified literature procedure.[174] To a solution of *N*-hydroxyphthalimide (20 g, 122.5 mmol) in DMF (120 mL), DBU (18.3 mL, 122.5 mmol) was added dropwise with stirring to give a very deep red solution. To this solution was added 1,2-dibromoethane (5.37 mL, 61.2 mmol) and the mixture was heated at 85 °C for 1 h during which time the deep red color faded to produce a colorless solution. The resulting solution was poured into ice and the resulting precipitate was filtered and washed with cold H_2O (50 mL) followed by cold CH_3CN (30 mL). The crude 1,2-diphthalimidooxyethane (**S1**) was recrystallized from n-butanol (3 L). A suspension of 1,2-diphthalimideooxyethane (16.0 g, 41.6 mmol) in glacial acetic acid/HCl (50 mL, 30:20, v/v) was heated at 115 °C for 3 h to give a clear solution. All the solvent and reagents were removed *in vacuo*. H_2O (10 mL) was added to the solid and the suspension was filtered and washed with HCl (6 M). The combined filtrate was collected and reduced to dryness. The crude product was recrystallized from EtOH/H_2O (5:1, v/v) to give the compound **1** ($C_2H_8N_2O_2$·2HCl salt) as white crystals (3.7 g, 22.4 mmol, 37 % over two steps and two recrystallizations); mp 228 °C; ^{1}H NMR (δ, DMSO): 11.16 (s, 6H), 4.26 (s, 4H). ^{13}C NMR (δ, DMSO): 72.21.

N-(2-(aminooxy)ethoxy)-5-(dimethylamino)naphthalene-1-sulfonamide (1).



To a solution of **S2** (61 mg, 0.37 mmol) in CH_2Cl_2 (3 mL), TEA (123 µL, 0.87 mmol) and DMAP (5 mg, 0.04 mmol) were added with stirring to give a clear solution. To this solution was added dansyl chloride (20 mg, 0.074 mmol) and the mixture was stirred at rt for 3 h. The solvent was removed *in vacuo* and the crude product was further purified by silica gel column chromatography with gradient elution (EtOAc:MeOH in 0.5% TEA) from 1:0 (v/v) going to 9:1 (v/v) to afford 14.2 mg of compound **1** (0.044 mmol, 59% yield) as a yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.600 (d, J = 8.0 Hz, 1H), 8.342 (d, J = 7.5 Hz, 1H), 8.298 (d, J = 9.0 Hz, 1H), 7.554 (m, 2H), 7.180 (d, J = 7.5 Hz, 1H), 4.083 (t, J = 4.5 Hz, 2H), 3.774 (t, J = 4.5 Hz,1H), 2.886 (s, 6H), ^{13}C NMR (125 MHz, $CDCl_3$) δ 132.18, 131.73, 128.85, 123.23, 118.18, 115.31, 75.03, 72.56, 45.4.

Kinetic analysis of oxime ligation reaction: rate constant analysis.

Dawson and coworkers originally established that the kinetics of oxime ligation reactions catalyzed by aniline fit a second order model that is first order in both aldehyde and alkoxyamine.[142] They also showed that the apparent second order rate constant for this process varied with aniline concentration. Wen-jun et al extended those observations by demonstrating that the apparent second order rate constant varied linearly with aniline.[161] Accordingly, we analyzed the rate data for the model reactions described here and confirmed that both the aniline- and mPDA-catalyzed reactions vary linearly with

catalyst concentration. In order to obtain an equation to curve fit the data, we started with the previously reported rate equation for the aniline catalyzed oxime reaction[3]:

Eq. S1 $V=k_{obs}\bullet[aldehyde][aminooxy]$

Where V is the rate of the reaction which is the rate of formation of the oxime as the end product, and k_{obs} is the observed rate constant of the reaction.

In general, for a second order reaction between A and B as reactants, the equation below for the kinetics of the reaction can be derived:

Eq. S2 $$\frac{[A]}{[B]} = \frac{[A]_0}{[B]_0} e^{([A]_0-[B]_0)kt}$$

By substituting [A] with $\{[Aminooxy]_o\text{-}[Oxime]\}$ and [B] with $\{[Aldehyde]_o\text{-}[Oxime]\}$ and solving the equation for [Oxime] the following relationship is obtained:

Eq. S3 $$[Oxime]=\frac{[Aminooxy]_0*[e^{([Aminooxy]_0-[Aldehyde]_0)kt}-1]}{\frac{[Aminooxy]_0}{[Aldehyde]_0}[e^{([Aminooxy]_0-[Aldehyde]_0)kt}]-1}$$

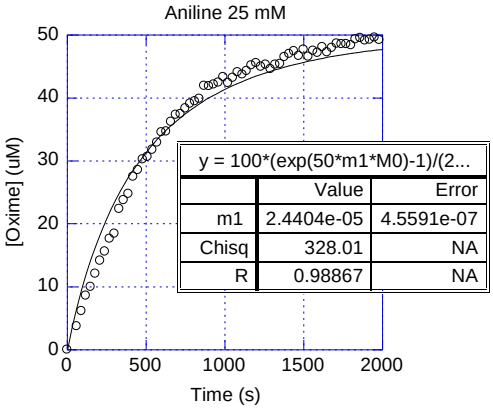
where [Oxime] is the concentration of product, $[Aminooxy]_o$ is the initial concentration of aminooxy-dansyl **1** and $[Aldehyde]_o$ is the initial concentration of aldehyde.

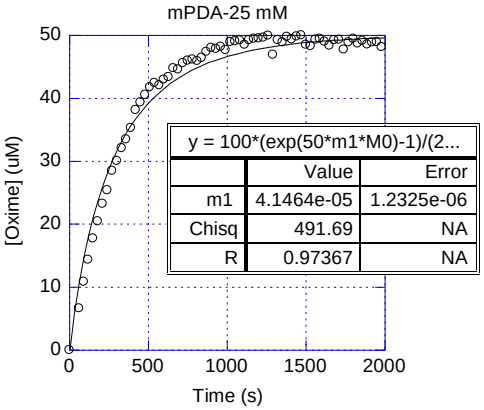
Fluorescence assay data analysis. Oxime-forming reactions were monitored by recording the increase in dansyl group fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) over time that occurred upon product formation. Hence, the raw experimental data was obtained in "fluorescence intensity/time" units. In order to obtain that data in units of "μM/time", it was necessary to convert the fluorescence values to concentrations. That was accomplished by first calculating the difference between the fluorescence intensity of the product and the starting reagents. Assuming complete reaction (100% conversion), that difference corresponds only to the fluorescence of the total amount of product. Thus, the

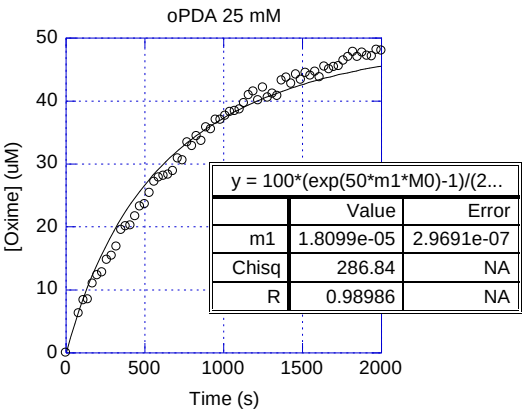
fluorescence of the starting reagents was subtracted from the raw fluorescence data followed by dividing by the fluorescence difference between the final product and starting material and then multiplied by the total concentration of the limiting reagent (in μM) which then gives the data in units of "μM/time". That calculation is summarized in equation S4:

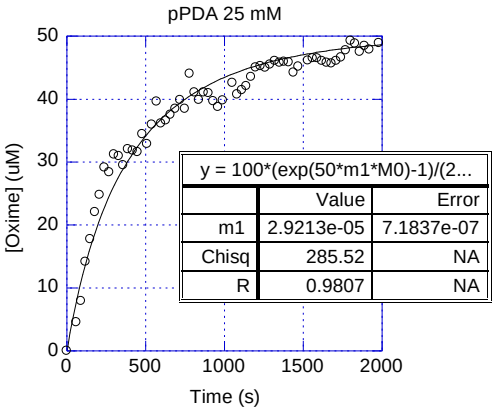
Eq. S4 $$[Oxime]_t = \frac{[F]_t - [F]_0}{[F]_{max} - [F]_0} \times [limiting\ reagent]_0$$

In that expression, $[Oxime]_t$ is the concentration of product at each given time, $[F]_t$ is the fluorescence of the reaction mixture at each given time, $[F]_0$ is the initial fluorescence of starting reaction mixture, $[F]_{max}$ is the maximum fluorescence of the reaction mixture at the end of the reaction and $[limiting\ reagent]_0$ is the initial concentration of the limiting reagent.


Aniline 25 mM
[Oxime] (uM)
Time (s)
y = 100*(exp(50*m1*M0)-1)/(2...
Value Error
m1 2.4404e-05 4.5591e-07
Chisq 328.01 NA
R 0.98867 NA


mPDA-25 mM
[Oxime] (uM)
Time (s)
y = 100*(exp(50*m1*M0)-1)/(2...
Value Error
m1 4.1464e-05 1.2325e-06
Chisq 491.69 NA
R 0.97367 NA


oPDA 25 mM
[Oxime] (uM)
Time (s)
y = 100*(exp(50*m1*M0)-1)/(2...
Value Error
m1 1.8099e-05 2.9691e-07
Chisq 286.84 NA
R 0.98986 NA


pPDA 25 mM
[Oxime] (uM)
Time (s)
y = 100*(exp(50*m1*M0)-1)/(2...
Value Error
m1 2.9213e-05 7.1837e-07
Chisq 285.52 NA
R 0.9807 NA

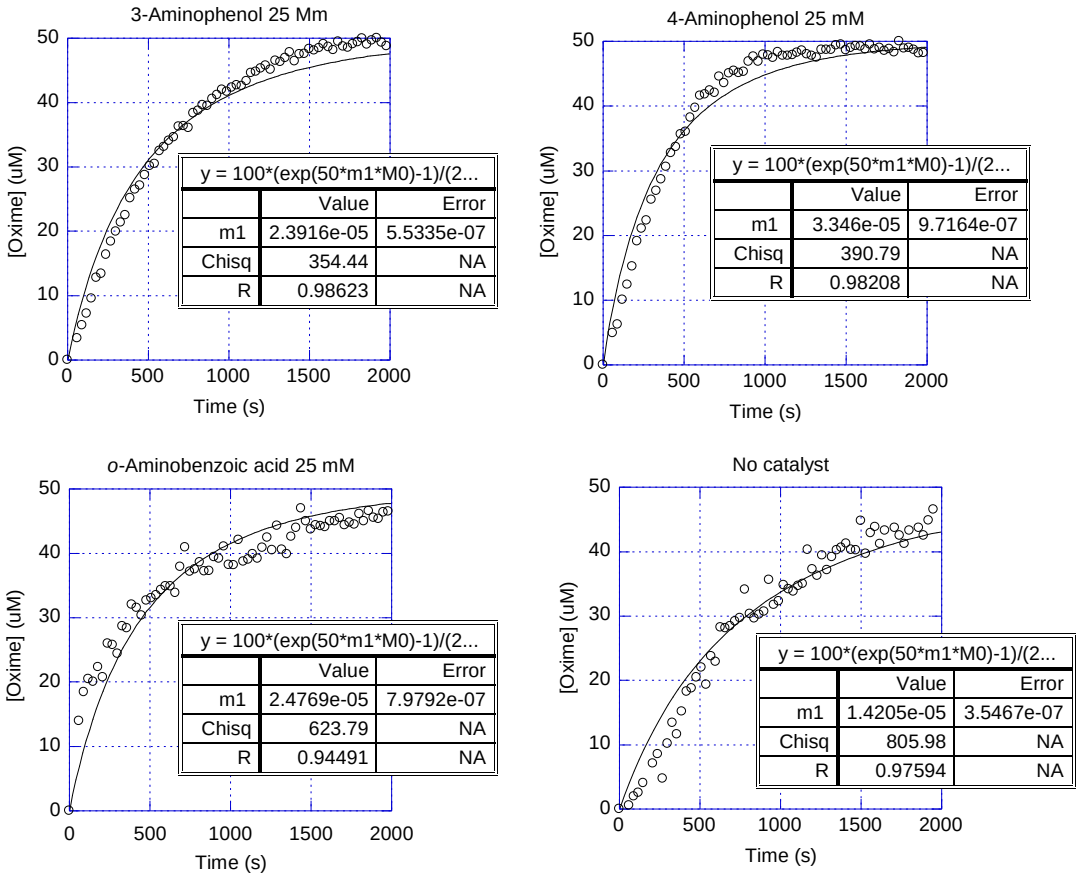


Figure 4.S1. Kinetic analysis of oxime ligation reaction, using 50 μM aldehyde (citral), 100 μM aminooxy-dansyl **1**, and 25 mM of catalysts. The k_{obs} values were obtained by fitting the experimental data to equation S3.

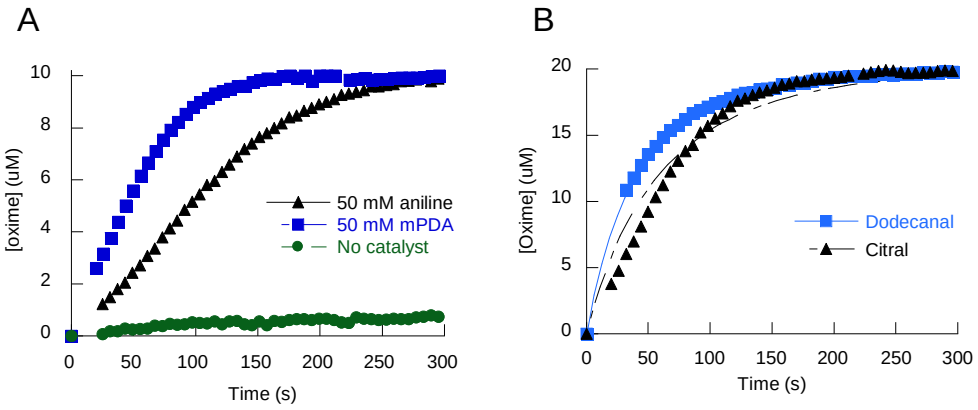


Figure 4.S2. A) Analysis of the reaction of aminooxy-dansyl **1** (10 μM) with citral (50 μM), a conjugated hydrophobic aldehyde, in presence of either 50 mM mPDA (squares) or 50 mM aniline (triangles) or no catalyst (circles). B) Kinetic analysis of oxime ligation reaction, using 50 μM aldehyde (citral or dodecanal), 20 μM aminooxy-dansyl **1** and 50 mM mPDA.

Kinetic analysis of oxime ligation between 2-pentanone and aminooxy-dansyl 1.

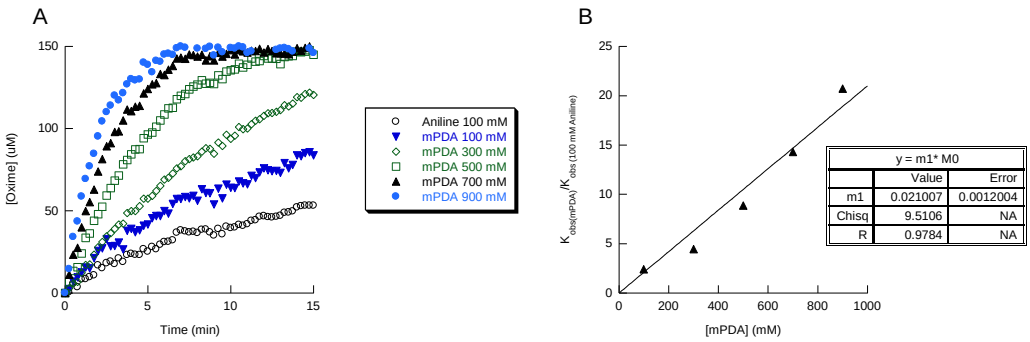
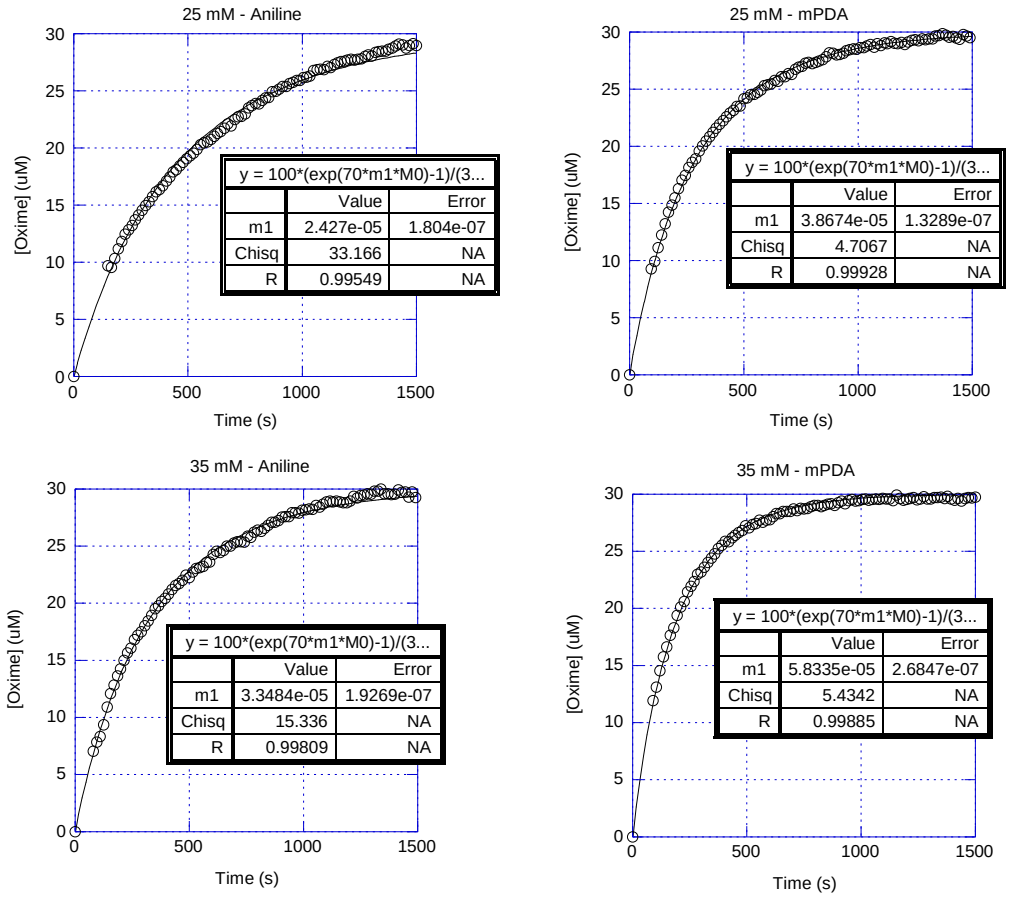


Figure 4.S3. A) Kinetic analysis of oxime ligation between a ketone (2-pentanone) and aminooxy-dansyl **1**. Reaction mixtures contained Tris·HCl (50 mM, pH 7.5), aminooxy-dansyl **1** (150 µM), 0.4 % (*w/v*) *n*-dodecyl-*ß*-D-maltoside, 5 mM ketone (2-pentanone) and either 100 mM aniline or varying concentrations of mPDA, in a final volume of 200 µL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the aminooxy, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 4 h. B) Plot of ratio of rate constant of mPDA catalyzed oxime ligation over aniline (100 mM) catalyzed oxime reaction vs concentration of mPDA used. The k_{obs} values were obtained by fitting the experimental data to equation S3.

Effect of the catalyst concentration on the kobs. In order to analyze the effect of the catalyst concentration on the k_{obs}, reactions with different catalyst concentrations were performed (Figure 4.S4). After analyzing the kinetics of the reactions and curve-fitting the data to the second order kinetic equation described above (Eq. S3), we observed that in the cases of both aniline and mPDA as the catalyst, a good fit was obtained to the kinetic expression derived above. Additional analysis revealed that reaction rate is linear with respect to the catalyst concentration (Figure 4.S5) which is in agreement with previous reports.[161]



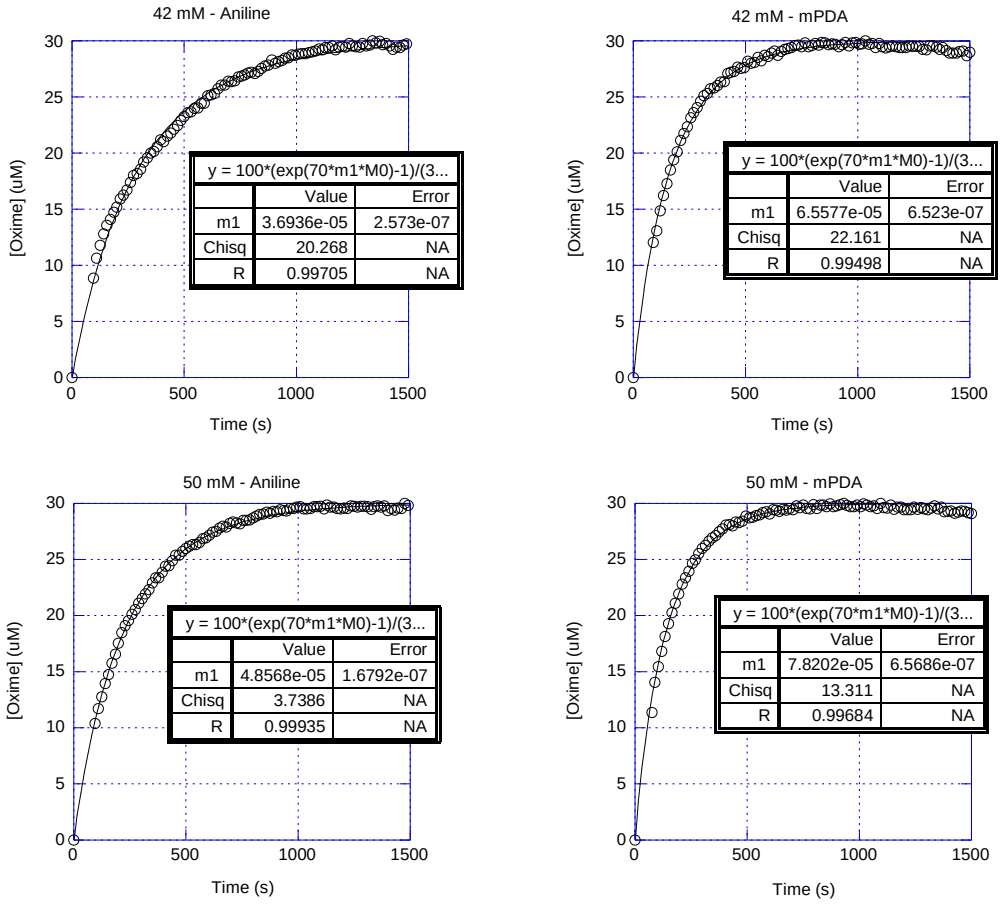


Figure 4.S4. Kinetic analysis of oxime ligation reaction, using 30 μM aldehyde (citral), 100 μM aminooxy-dansyl **1**, varying concentrations of either *m*-phenylenediamine or aniline (from 25 to 50 μM). The data was fit to the equation S3.

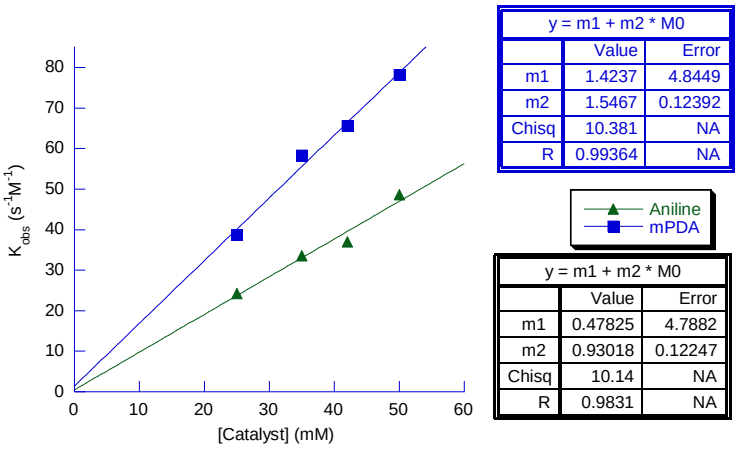


Figure 4.S5. Kinetic analysis of oxime ligation reaction to determine the relationship between the observed rate constant and the catalyst concentration. Experiments were performed using 30 μM aldehyde (citral), 100 μM aminooxy-dansyl **1** and varying concentrations of either *m*-phenylenediamine or aniline (from 25 to 50 μM).

GFP-CVIA preparation. Protein was prepared as previously described with one modification.[136] In the final phenyl sepharose chromatography step, after the protein was loaded onto the column and washed with buffer as explained in the original work, the protein was eluted from column by adding water instead of buffer.

LC-MS analysis of proteins for determination of prenylation and labeling efficiency. Purified prenylated GFP (**4a**) and pure GFP-CVIA (**3**) were analyzed by LC-MS to ensure complete prenylation. Proteins were stored in Tris·HCl (50 mM, pH 7.5) prior to injection into the LC-MS instrument. Crude reaction mixtures of GFP-aldehyde **4a** and aminooxy **1** catalyzed by either aniline or *m*-phenylenediamine were analyzed by LC-MS to ensure complete ligation in both cases of the catalysts. The LC-MS method used was a gradient of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

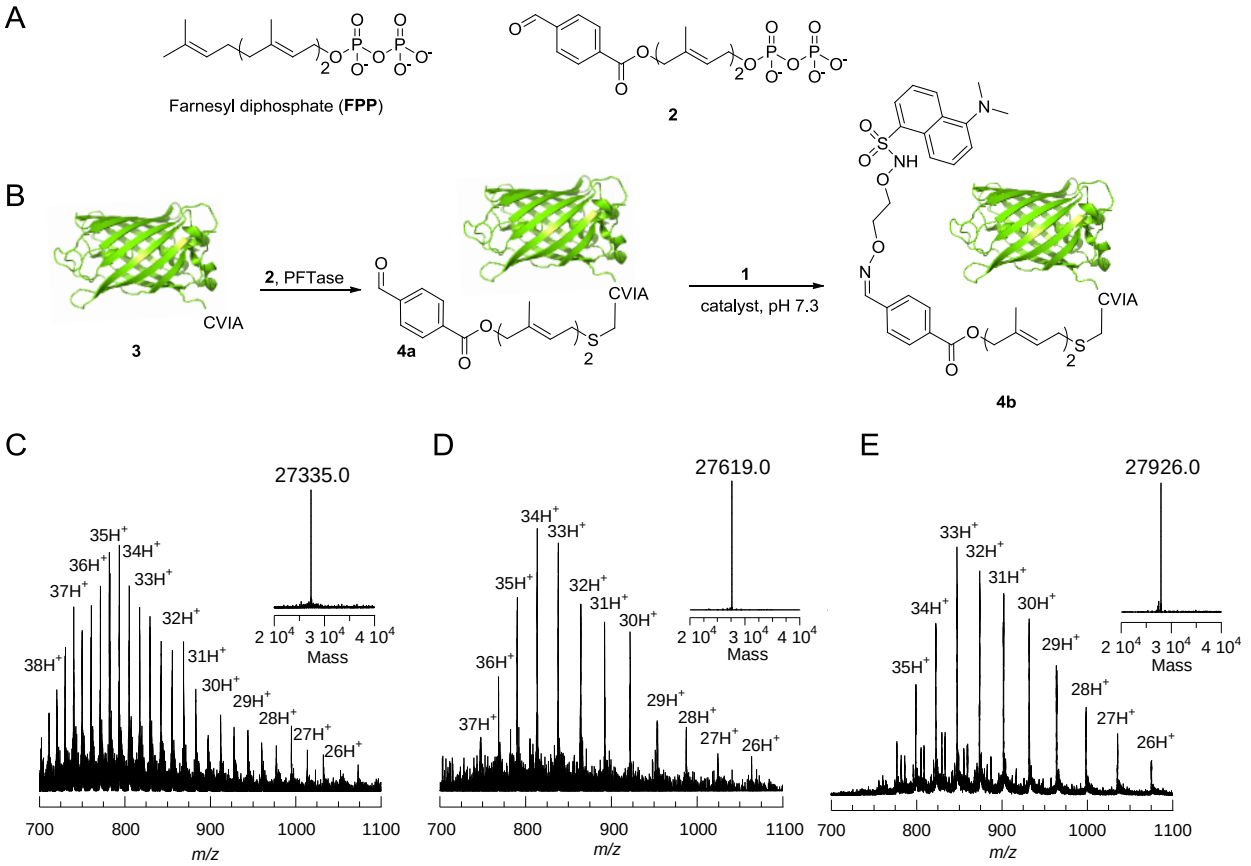


Figure 4.S6. A) Structures of farnesyl diphosphate (**FPP**) and formylbenzoyl-oxy geranyl diphosphate (**2**). B) Schematic representation of prenylation of GFP-CVIA **3** with aldehyde-containing analog **2** to yield the prenylated product **4a** and subsequent oxime ligation with **1** to yield oxime **4b**. C, D and E) ESI MS analysis of **3**, **4a** and **4b** with the deconvoluted mass spectra shown in the insets, respectively.

Kinetic analysis of protein labeling via oxime ligation. Reaction mixtures contained PB (100 mM, pH 7.0), 10 μM GFP-aldehyde **3**, 50 μM aminooxy-dansyl **1** and varying concentrations of *m*-phenylenediamine or 100 mM aniline, in a final volume of 250 μL. The reaction mixtures were equilibrated at rt for 1 min, initiated by the addition of the catalyst, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 100 min.

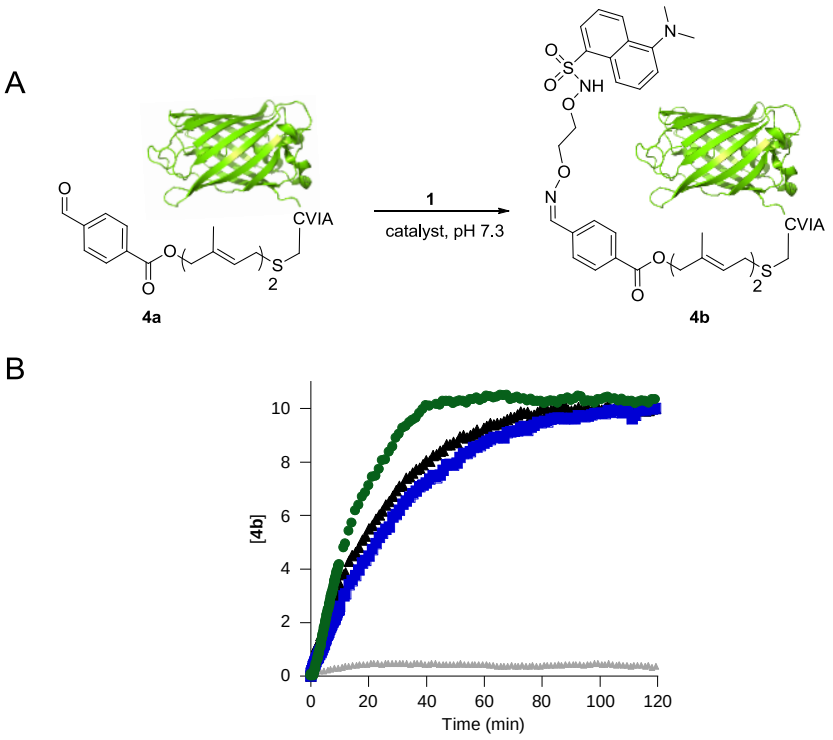


Figure 4.S7. Schematic representation of labeling of GFP-aldehyde **4a** with aminooxy **1** to yield oxime **4b**. B) Kinetic analysis of oxime ligation reaction between **4a** (10 µM) and **1** (50 µM) using 50 mM *m*-phenylenediamine (green circles), 50 mM aniline (blue squares) or no catalyst (black triangles). Gray triangles: GFP-CVIA **3** (10 µM, with no aldehyde attached to it) is treated with aminooxy **1** (50 µM) in presence of 50 mM *m*-phenylenediamine to further confirm that reaction is truly bioorthogonal.

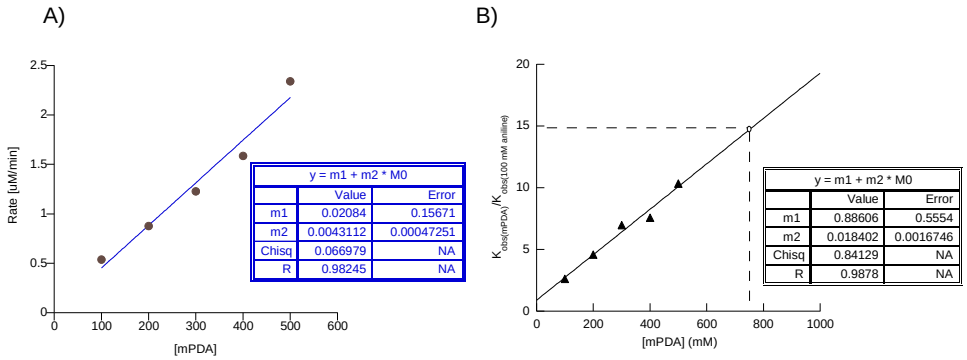


Figure 4.S8. A) Kinetic analysis of oxime ligation reaction between aminooxy **1** and aldehyde-GFP **4a** to determine the relationship between the observed initial rate of the reaction and the catalyst concentration. Experiments were performed using 10 μM aldehyde-GFP **4a**, 50 μM aminooxy-dansyl **1** and varying concentrations of *m*-phenylenediamine. In this case when higher concentrations of the catalyst (>500 mM) were used (where the ratio [cat]/[aminooxy] was high), the Schiff base between the catalyst and aldehyde becomes a significant product, complicating the kinetic analysis; hence that data was omitted from this analysis. It should be noted if it is necessary to employ very high catalyst concentrations, Schiff base formation can be countered by raising concentration of the aminooxy reagent. B) Plot of ratio of rate constant of mPDA catalyzed oxime ligation over aniline (100 mM) catalyzed oxime reaction vs concentration of mPDA used. Extrapolation of the line shows that at 750 mM of mPDA reaction would be 15 times more efficient in case of mPDA relative to 100 mM aniline.

Table 4.S1. Kinetic analysis of oxime ligation reaction between aldehyde-GFP **4a** and aminooxy-dansyl **1**, using aniline or mPDA as catalysts.

[Catalyst][a] mM	Initial rate ($\mu M\ s^{-1}$)	$\frac{\text{Rate (x mM mPDA)}}{\text{Rate (100 mM aniline)}}$	k_{obs} [b] ($s^{-1}\ M^{-1}$)	$k_{obs(\text{mPDA})}/k_{obs(\text{aniline})}$	$t_{1/2}$ (min)
100 aniline	0.226	-	10.35±0.05	-	23.7
100 mPDA	0.539	2.38	27.04±0.33	2.61	9.06
200 mPDA	0.878	3.87	48.45±0.74	4.68	5.18
300 mPDA	1.228	5.42	71.84±1.53	6.94	3.41
400 mPDA	1.586	7.00	78.11±1.01	7.54	3.14
500 mPDA	2.343	10.34	107.36±2.13	10.37	2.3

[a]Reactions were performed using 10 μM aldehyde-GFP **4a**, 50 μM aminooxy-dansyl **1**, and the catalyst concentrations given.
[b]The k_{obs} values were obtained by fitting the experimental data to Equation S3 using Kaleidagraph v4.1.3. The values are provided ± the standard error obtained from the curve fit.

Table 4.S2. Kinetic analysis of the release of the hydrazone immobilized protein into fluorescently labeled oxime protein via hydrazone-oxime exchange reaction.

[Catalyst][a] mM	k_{obs} [b] (h^{-1})	$t_{1/2}$ (h)	$k_{obs(\text{mPDA})}/k_{obs(\text{aniline})}$
100 aniline	0.016±0.001	43.6	-
750 mPDA	0.237±0.009	2.9	15

[a]Immobilized protein was incubated with aminooxy fluorophore **5** (1 mM) and catalyst, followed by analysis of the amount of released protein in the solution via SDS-PAGE.
[b]The k_{obs} values were obtained by fitting the experimental data to a pseudo first order reaction. The values are provided ± the standard error obtained from the curve fit.

CNTF-CVIA preparation. The gene for Ciliary Neurotrophic Factor (CNTF) was purchased from DNA 2.0 on a pJexpress414 vector. The sequence of this synthetic gene (Menlo Park, CA) is given below. The CVIA portion of the protein was added using an Invitrogen QuikChange Site Directed Mutagenesis kit (catalogue #200523 Menlo Park, CA) following the manufacturer's instructions.

The forward primer was tatggtgcaaaagataaacaaatgtgcgtgattgcgtaactcgagccccctag, and the reverse primer was ctaggggctcgagttacgcaatcacgcacatttgtttatcttttgcaccata. A plasmid containing the CNTF-CVIA was transformed into BL21(DE3)pLysS *E. coli* cells.

BL21(DE3)pLysS *E. coli* cells containing the CNTF-CVIA plasmid were plated on LB-Agar plates containing 100 μg/mL ampicillin. These plates were grown overnight at 37 ^{o}C. Single colonies were then picked and used to innoculate 50 mL of LB media containing 100 μg/mL ampicillin. These flasks were grown overnight with shaking at 250 rpm at 37 ^{o}C. 10 mL of the overnight growth was added to 1 L of LB media containing 100 μg/mL of ampicillin and incubated at 37 ^{o}C with shaking at 250 rpm. This culture was grown to an OD_{600} of 0.8 at which time protein expression was induced by the addition of 1 mL of 1M IPTG. Cultures were then incubated for an additional 4 h by shaking at 250 rpm at 37 ^{o}C followed by harvesting by centrifugation at 5,400g for 10 min. *E. coli* cell pellets were stored at -80 ^{o}C. A cell pellet corresponding to 1 L of cell growth was suspended in 50 mL of buffer containing 50 mM Tris•HCl, pH 7.5 and 5 mM 2-mercaptoethanol. This was then subjected to pulse sonication (10 sec sonication / 10 sec off) at 50 W for a total sonication time of 5 min. The sonicated solution was then centrifuged at 13,000g for 30 min. The supernatant was removed and the insoluble protein pellet containing the CNTF-CVIA was suspended in 30 mL of buffer containing 10 mM Tris•HCl, pH 7.5, 6 M guanidine•HCl, 5 mM 2-mercaptoethanol, 100 mM NaH_2PO_4, and 20 mM imidiazole. To thoroughly suspend the protein, the solution was subjected again to pulse sonication (10 sec sonication / 10 sec off) at 50 W for a total sonication time of 5 min. This solution was then added at 3 mL/h to 500 mL of refolding

buffer (50 mM Tris•HCl, pH 7.5, 0.5 M NaCl, 10 mM CHAPS, 2 mM DTT). This solution was then left to stir overnight at 4 ^{o}C and then concentrated using an Amicon Centriprep centrifugation (10,000 MW cut-off) according to manufacturer's instructions. This yielded 9.2 mL of a 6.5 mg/mL protein solution. This solution was diluted in half to store in 40% glycerol at -80 ^{o}C.

CNTF gene from DNA 2.0 containing N-terminal His-tag.

Protein sequence:

MHHHHHHLVP • RGSMAFAEQT • PLTLHRRDLC • SRSIWLARKI •

RSDLTALMES • YVKHQGLNKN • INLDSVDGVP • VASTDRWSEM •

TEAERLQENL • QAYRTFQGML • TKLLEDQRVH • FTPTEGDFHQ •

AIHTLMLQVS • AFAYQLEELM • VLLEQKIPEN • EADGMPATVG •

DGGLFEKKLW • GLKVLQELSQ • WTVRSIHDLR • VISSHQMGIS •

ALESHYGAKD • KQM

DNA sequence (with ATG start underlined):

aggagatatctaga<u>atg</u>caccatcatcatcaccacctggttccacgcggtagcatggcc
ttc
E I S R M H H H H H H L V P R G S M A
F
gctgaacaaaccccgctgacgctgcaccgtcgcgatctgtgctcccgtagcatctggct
g
A E Q T P L T L H R R D L C S R S I W L
gcccgcaagattcgtagcgacctgaccgcattgatggaatcttacgttaagcatcaagg
t
A R K I R S D L T A L M E S Y V K H Q G
ctgaacaaaaacattaatctggatagcgtggatggtgttccggtcgcgagcacggaccg
t
L N K N I N L D S V D G V P V A S T D R
tggagcgaaatgaccgaagcggagcgcctgcaggagaacctgcaggcatatcgtacctt
c
W S E M T E A E R L Q E N L Q A Y R T F

caaggtatgctgaccaaactgctggaggatcaacgcgtgcactttacgccgaccgaaggt
 Q G M L T K L L E D Q R V H F T P T E G
gattttcatcaggcgatccacaccctgatgctgcaagttagcgcttttgcttaccagctg
 D F H Q A I H T L M L Q V S A F A Y Q L
gaagagctgatggtgttgttggaacagaagattccggagaatgaggccgacggtatgccg
 E E L M V L L E Q K I P E N E A D G M P
gcgaccgtcggcgacggtggcctgttcgaaaagaagctgtggggcctgaaagttctgcag
 A T V G D G G L F E K K L W G L K V L Q
gagctgagccagtggacggtccgttccattcatgacctgcgtgtgattagcagccaccaa
 E L S Q W T V R S I H D L R V I S S H Q
atgggtatcagcgcactggaatctcattatggtgcaaaagataaacaaatgtaactcgag
 M G I S A L E S H Y G A K D K Q M - L E

Primers for CVIA mutant construction (CVIA sequences underlined)**.**

Forward Primer
5′tatggtgcaaaagataaacaaatgtgcgtgattgcgtaactcgagccccctag-3′
Translated 5′-3′ Frame 1
tatggtgcaaaagataaacaaatgtgcgtgattgcgtaactcgagccccctag
 Y G A K D K Q M C V I A - L E P P
Reverse Primer
5′-ctaggggctcgagttacgcaatcacgcacatttgtttatcttttgcaccata-3′
Translated 3′-5′ Frame 1
tatggtgcaaaagataaacaaatgtgcgtgattgcgtaactcgagccccctag
 Y G A K D K Q M C V I A - L E P P

Coupling reaction between aldehyde-labeled CNTF-CVIA (9) with alexafluor-488 (5). Alexafluor-488 (**5**) (4.2 µL of 3.2 mM solution in DMSO) was added to 42 µL of **9** (stock solution of 60 µM in Tris·HCl (50 mM, pH 7.5)). PB (2 M, pH 7, 2.5 µL) was added and the reaction was initiated by adding 50 mM *m*-phenylenediamine (stock solution of 1.5 M in 0.3 M PB, pH 7.0) and was allowed to proceed for 2 h at rt. LS-MS

analysis of the sample showed only oxime ligated protein and no free aldehyde was detected indicating a complete reaction in both prenylation and oxime ligation reactions.

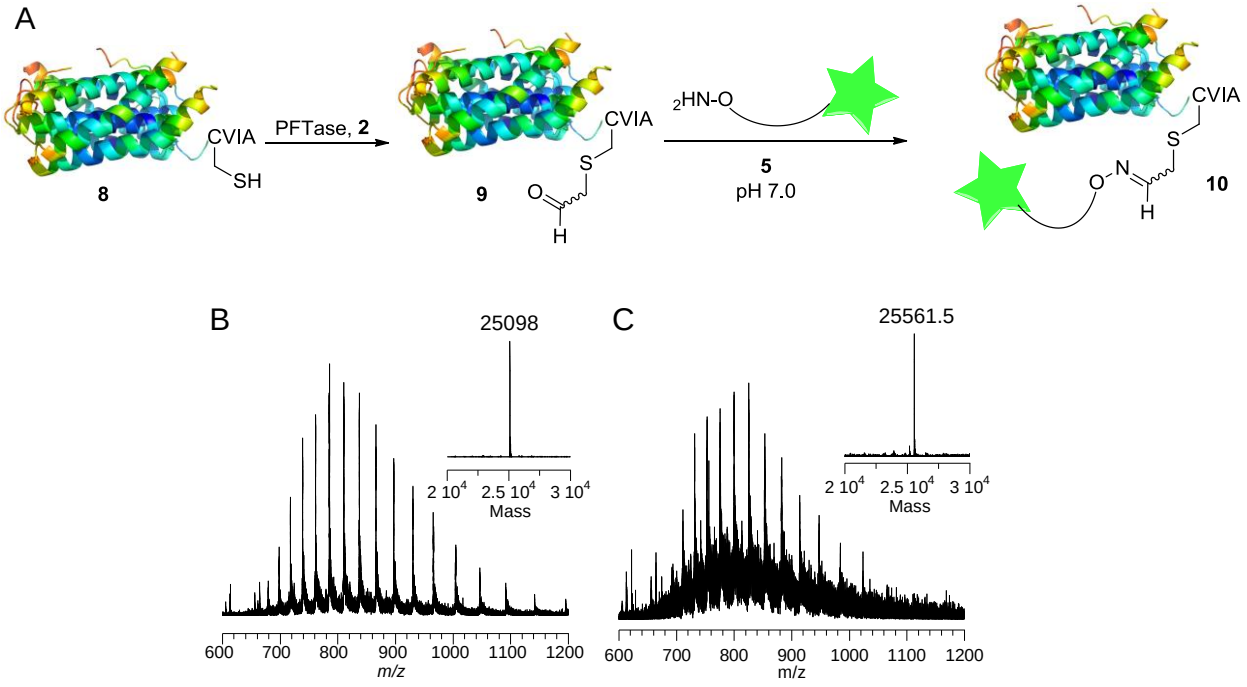


Figure 4.S9. A) Schematic representation of prenylation following by oxime ligation of **8** to yield **9** and **10** respectively. B and C) ESI MS spectra of prenylated-CNTF **9** (spectrum B) and oxime product **10** (spectrum C) with the deconvoluted mass spectra shown in the insets.

Preparation of DHFR2 M174pAcF (11). The unnatural amino acid *p*-Acetyl phenylalanine was synthesized and characterized as described previously.[175] The BL21 (DE3) competent cells were purchased from Invitrogen™. Experimental details of the plasmid encoding two cysteine free DHFR (DHFR^2) fusion proteins connected with 13 amino acid linker has been described previously.[176] The site for the unnatural amino acid incorporation (M174TAG) was chosen on the basis of the surface accessibility of the residue. The site was mutated to an amber (TAG) stop codon with a Quick change® site

directed mutagenesis kit (Strategene). The primers used for the mutation were RS_M174TAG F and RS_M174TAG R as shown below. To facilitate the isolation of the fully expressed DHFR[2] M174 pAcF protein, a C-terminus 6xHis sequence was appended. The primers used for the His_6 insertion were RS_M174TAG His_6 F & RS_M174TAG His_6 R as shown below. The plasmid pEVOL_pAcF encoding amino acyl-tRNA synthetase (MjTyrRS) and $tRNA_{CUA}$ evolved from *M. jannaschi* was provided by Dr. Schultz group and has been described previously.[177]

Primer	Seuquence
RS_M174TAG F	GGT GGT TAG GTT CCG CGT GGT
RS_M174TAG R	ACC ACG CGG AAC CTA ACC ACC
RS_M174TAG His_6 F	C GAA ATC CTC GAG CGT CGT TAG CAC CAC CAT CAC CAT CAT **TAA GGA** TCC TAA TTA ATT AAT TCA C
RS_M174TAG His_6 R	G CTA ACG ACG CTC GAG GAT TTC GAA ACT ATA GCT ATG CGA G

To express the DHFR[2] fusion protein with *p*-acetyl phenylalanine, the plasmid encoding DHFR[2] M174TAG was co-transformed with pEVOL_pAcF into BL21 (DE3) competent *E. coli* cells. 50 ml overnight cultures in LB (Luria-Bertani) media were used to inoculate 500 mL of M9 minimal media containing chloramphenicol (34 mg/ml) and ampicillin (50 mg/ml). Cultures were then grown at 37 °C until the O.D. reached 0.57 after which the

protein expression was induced by adding 0.3 mM IPTG, 0.04% arabinose and 1 mM pAcF. The cultures were then transferred to 30 °C and incubated with shaking for an additional 18 h. The cells were then harvested via centrifugation at 7,500 rpm for 10 min. The cells pellet was resuspended into lysis buffer containing 1 mg/ml of lysozyme, 20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole with gentle shaking for 30 min. The partially lysed cells were then cooled and sonicated. The lysate was then centrifuged at 16,000 rpm for 45 min and the supernatant was then loaded onto a Ni-NTA agarose column. The column was washed with buffer A (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole) and eluted with gradient buffer B (20 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole). The protein concentration and purity in the fractions was determined using Bradford assay and SDS PAGE, respectively. After pooling the fractions containing the desired protein, it was concentrated and exchanged into 0.1 M PB pH 7.0 using an Amicon ultrafilitration device equipped with a 30 kDa membrane. The final yield of the protein was 8 mg/L and was stored in - 80 °C. The production of the final $DHFR^2$ M174 pAcF protein was confirmed by ESI-MS (Calculated m/z 38061.5, found 38062.6).

PEGylation of DHFR2 M174pAcF (11) with aminooxy-PEG (12, 3 kDa) using mPDA. Aminooxy-PEG **12** (5 mM) was reacted with $DHFR^2$ M174pAcF **11** (10 μM) in PB (0.1 M, pH 7) in presence of either 100 mM aniline, 500 mM mPDA or no catalyst. The amounts of PEGylated protein in the solutions were analyzed at different times via SDS-PAGE. Gels were visualized by staining with Coomassie blue. Densitometry analysis on the gels was performed using the program ImageJ v1.46.

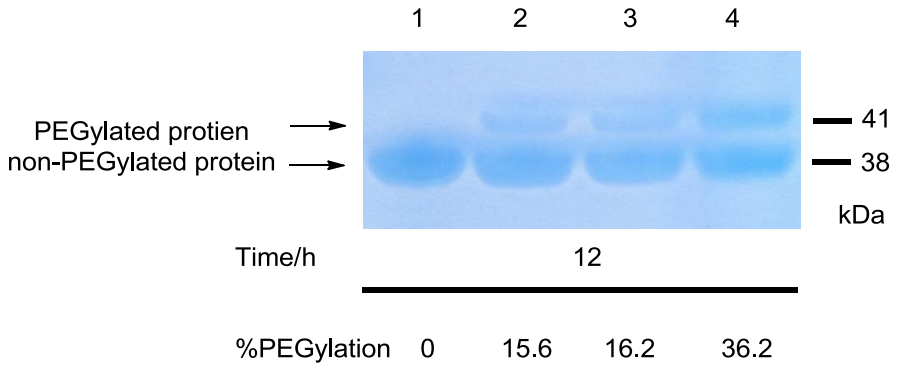
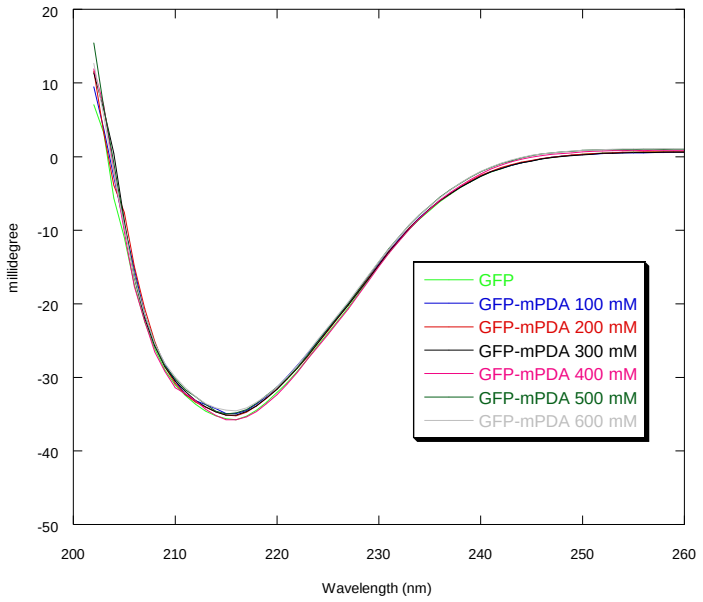


Figure 4.S10. SDS-PAGE analysis of PEGylation of the protein DHFR[2] M174pAcF which contains the unnatural amino acid *p*-Acetyl phenylalanine with aminooxy-PEG **12** (3 kDa). Protein (10 µM) was incubated with aminooxy-PEG (5 mM) with or without catalyst for 12 h. Next the samples were analyzed via SDS-PAGE. Lane 1: pure protein; lane 2: no catalyst was used; lane 3: 100 mM aniline was used as the catalyst; lane 4: 400 mM mPDA was used as the catalyst. The bands were visualized by staining with Coomassie Blue. %PEGylation were obtained by densitometric analysis of the SDS-PAGE using the Coomassie blue stained gel.

Circular dichroism spectroscopy studies for analysis of the effect of mPDA on protein structure and function.

GFP-CVIA **3** (from stock solution of 92 µM diluted to final concentration of 18 µM) was treated with different concentrations of mPDA for ~30 min following by gel filtration chromatography using Zeba spin desalting column according to the manufacturer's instruction (Thermo Scientific), to remove the mPDA catalyst. Circular dichroism spectroscopy of the resulting samples showed no significant differences suggesting that exposure to high concentrations of mPDA does not cause substantial denaturation or irreversible protein unfolding.

Recovered GFP solutions from Zeba spin desalting columns were concentrated down using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Assuming 100% recovery from Zeba spin desalting columns, concentrations were calculated based on the volume of recovered concentrated solutions: pure GFP (21 µM, 440 µL), GFP that has been treated with 100 mM mPDA (21.5 µM, 430 µL); GFP that has been treated with 200 mM mPDA (22 µM, 410 µL); GFP that has been treated with 300 mM mPDA (17 µM, 540 µL); GFP that has been treated with 400 mM mPDA (18 µM, 500 µL); GFP that has



been treated with 500 mM mPDA (17 µM, 540 µL); GFP that has been treated with 600 mM mPDA (17 µM, 540 µL). Spectra were normalized based on 21 µM as the concentration of GFP solutions.

Figure 4.S11. Circular dichroism spectroscopy studies for analysis of the effect of mPDA on protein structure and function. Spectra were normalized to an equal concentration of

GFP-CVIA **3** (21 μM). Green line: pure GFP; blue line: GFP that has been treated with 100 mM mPDA for 30 min; red line: GFP that has been treated with 200 mM mPDA for 30 min; black line: GFP that has been treated with 300 mM mPDA for 30 min; pink line: GFP that has been treated with 400 mM mPDA for 30 min; dark green line: GFP that has been treated with 500 mM mPDA for 30 min; gray line: GFP that has been treated with 600 mM mPDA for 30 min.

Effect of catalyst concentration on the enzyme activity.

PFTase, as a model enzyme, was incubated with varying concentrations of either of the two catalysts, aniline or mPDA, for ~15 min, in PB (0.3 M, pH 7.0). Next, 10 μL of each of the PFTase solutions was added to a solution containing Tris•HCl (50 mM , pH 7.5) , $MgCl_2$ (10 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM) and 2.0 μM *N*-dansyl-GCVIA in a final volume of 200 μL. The reaction mixtures were equilibrated at 30 °C for 5 min, initiated by the addition of FPP (10 μM), and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 25 min. The initial rates of formation of products were obtained as slopes in IU/s using least squares analysis. Corrections were applied to all the rate calculations based on the difference between the fluorescence intensity of the prenylated product and the starting peptide. Assuming 100% conversion, the difference corresponds only to the fluorescence of the total amount of the product. The slope was then divided by the fluorescence difference followed by multiplying by the total concentration of peptide (2.0 μM) which then gives the rate of formation of product in μM/s. The rates for the different reactions were normalized to the rate observed in the absence of any catalyst (set to 100%) to facilitate comparison.

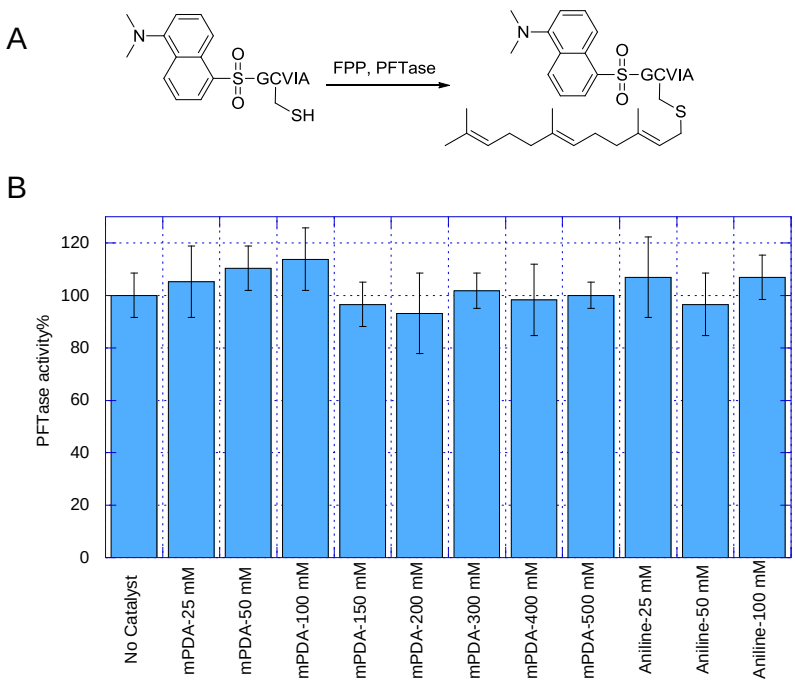


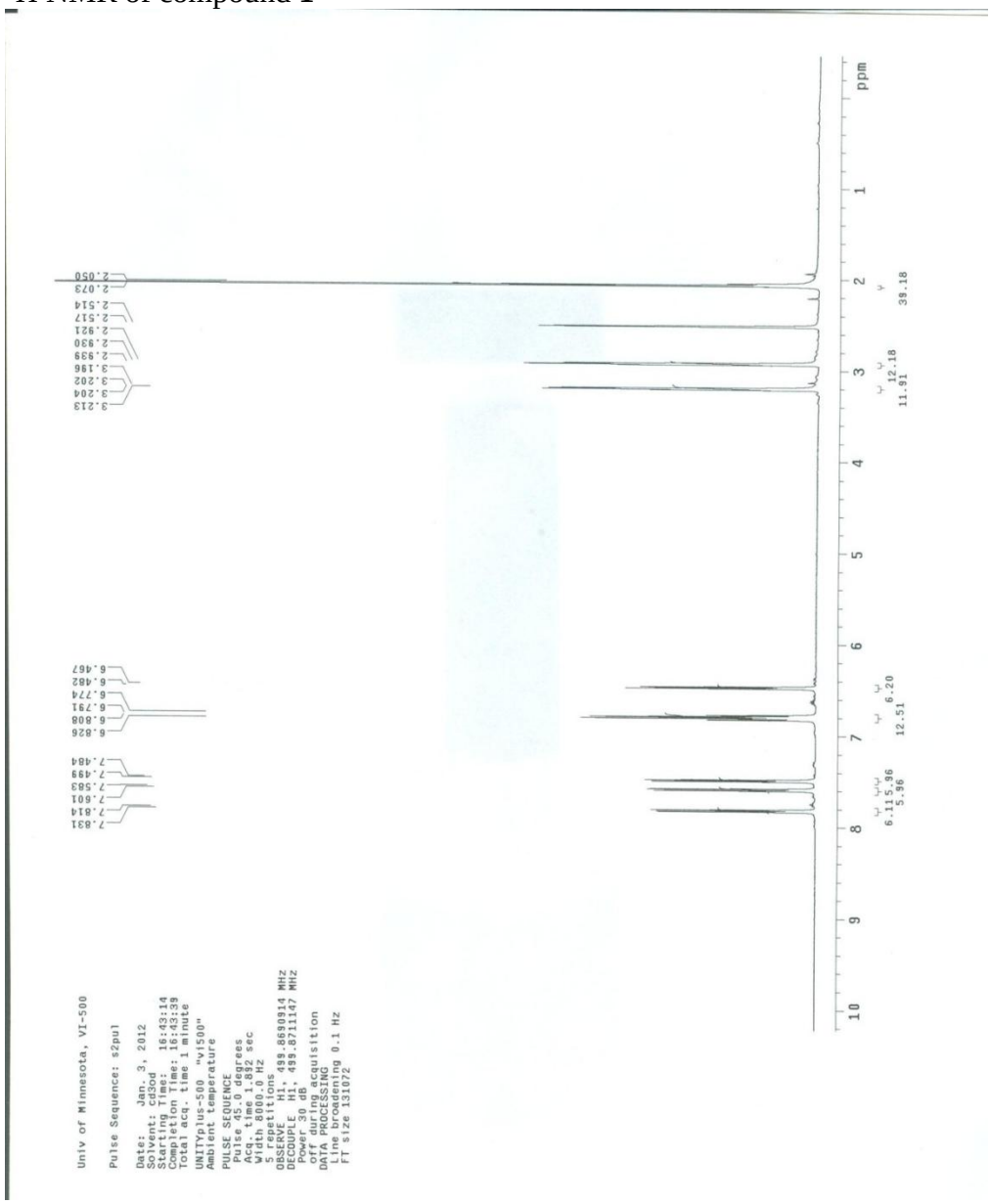
Figure 4.S12. Effect of catalysts on the activity of PFTase. A) Schematic representation of the farnesylation reaction of *N*-dansyl-GCVIA using PFTase. B) Rate analysis of the enzymatic farnesylation reaction using varying concentrations of either aniline or mPDA. The rates for the different reactions were normalized to the rate observed in the absence of any catalyst (set to 100%) to facilitate comparison. Reactions were performed in duplicates.

General procedure for MALDI analysis of protein samples. The sample was adsorbed onto a zip-tip (C_4 column) via repeated cycles of aspiration and ejection (5-10 cycles of 10 µL each) using a pipettor. Next, in order to remove excess buffer and reagents, the zip-tip was washed 5x10 µL with solvent A (H_2O containing 0.1% TFA; v/v) and the proteins eluted with 2 µL of a mixture of A and B (25:75) (solvent B: CH_3CN containing 0.1% TFA; v/v). Next 0.7 µL of the eluted material was added to a MALDI plate and 0.7 µL of matrix was added on top of the sample plate and both were mixed thoroughly

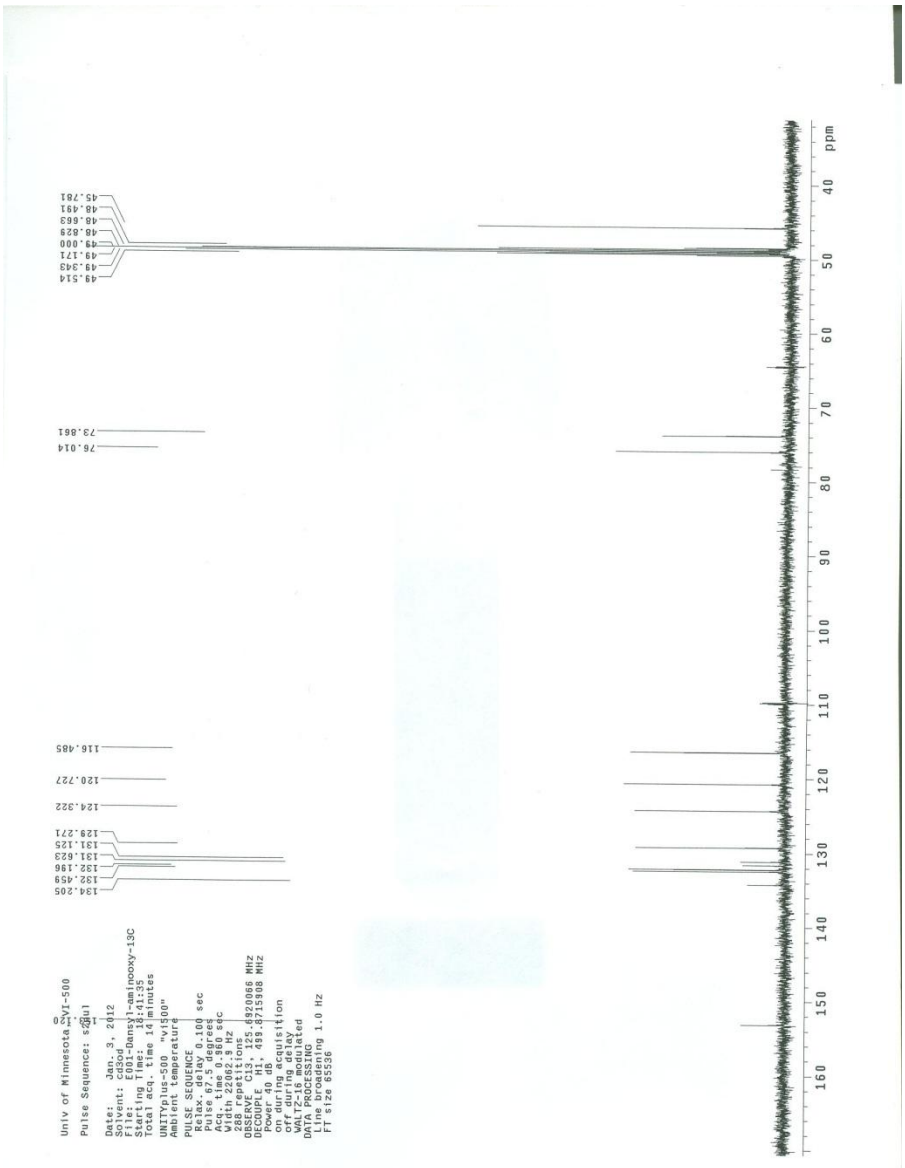
to form crystals. Saturated solution of sinapinic acid (3,5-dimeth-oxy-4-hydroxy-cinnamic acid) was used as the matrix.

NMRs.

^{1}H NMR of compound **1**



^{13}C NMR of compound **1**.



Chapter 5. Macromolecular Protein Self-Assembly by Bioorthogonal Chemical Dimerization

Mohammad Rashidian, Sidath C. Kumarapperuma, Kari Gabrielse, Adrian Fegan, Mark D. Distefano and Carston R. Wagner

Construction of heterofunctional proteins is emerging as an interesting and important tool in biopharmaceuticals and in fact there are examples like immunoglobulin G Fc domain fusions that are among the growing class and top-selling human biotherapeutics. Combining a protein with other proteins or molecules, such as a targeting element, a toxic protein or molecule, a fluorophore or PEG groups, can improve the specificity, functionality, potency, and pharmacokinetic profile of the protein construct. Herein, we have demonstrated that enzymatic protein modification and bis-MTX driven chemical dimerization can be used to prepare complex multi-protein assemblies. Modular nature of this method can reduce the problems associated with heterogeneity of the conjugates and add to the versatility of the technique. This strategy is demonstrated with a scFv as the targeting element, GFP molecule representing the effector (eg. toxin) and a small molecule dye to represent a secondary effector (cytotoxic drug or an imaging agent). An important feature of this method is that a protein can be simultaneously and site-specifically labeled with two bioorthogonal functionalities.

Introduction. Self-organization of protein molecules into higher order structures is an integral part of living cells. A vast array of essential cellular functions are carried out through well-orchestrated non-covalent protein assembly and disassembly. By incorporating unique evolutionarily selected structural features, the formation of higher order structures by self-assembly are programmed by nature at the genetic level.[178–181] In principle, protein assemblies can be genetically engineered to form designed biomimetic macromolecular nanostructures [182,183] However, unlike DNA and RNA based

nanostructures,[184–187] the rules governing protein self-assembly are only beginning to be defined. Much of the previous effort to construct biomimetic protein assemblies has been based on direct protein-protein interactions through the incorporation of complementary amino acid residue surface interactions.[188] As an alternative, chemical biological approaches for macromolecular protein assembly based on metal coordination,[189] polymer conjugation[190], enzyme-inhibitor interactions,[191] supramolecular chemistry [192,193] and chemically induced dimerization[194] have begun to be explored.

Over the course of the last decade, bioorthogonal chemical methods have been developed for the site-specific chemical modification of proteins. These approaches have been used to enhance the functionality of proteins.[195,196] For example, fluorophores can be site-specifically attached to proteins as a biophysical or cellular localization tool, while protein-polymer conjugation is a well-established method for modulating the *in vivo* behavior of proteins.[49,118,119,121] Recently, employing non-natural mutagenesis techniques, Schultz and co-workers have coupled two antibody FABs via an alkyne-azide cycloaddition click reaction.[167] Bertozzi and coworkers, used formyl generating enzyme strategy to incorporate aldehyde moiety into proteins, then using bifunctional linkers that contain aminooxy and cyclooctyne or azide functionalities, they could synthesize heterobifunctional protein fusions.[197] In other work, Ploegh and coworkers showed that N-to-N and C-to-C protein conjugates could be made using a variation of sortagging. They first incorporated either an azide or alkyne, to the N-or C-termini of proteins using sortase strategy. Next, they reacted the two complimentary handles to prepare the desired dimer.[72]

Herein, we describe an alternative modular method for protein self-assembly that combines bioorthogonal protein labeling via enzymatic farnesylation and chemically induced dimerization for the preparation of GFP-labeled single chain antibody chemically self-assembled nanostructures (CSANs) (Figure 5.1). An important advantage of our methodology is the fact that a protein can be enzymatically simultaneously and site-specifically labeled with two bioorthogonal functionalities, an aldehyde and an alkyne group.

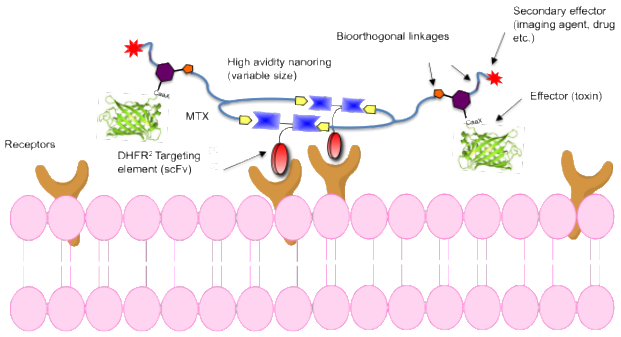


Figure 5.1. Modular construction of high avidity "effector-antibody-fluorophore" conjugates for therapeutic cargo delivery and tracking.

Harnessing the reversible high affinity binding of methotrexate (MTX) with dihydrofolate reductase (DHFR), we have demonstrated that recombinantly produced antiCD3 single-chain variable fragments (scFv) containing dimeric dihydrofolate reductase fusion proteins (**DHFR^{2}antiCD3)** can be induced to form macrocyclic rings with a bis-MTX dimerizer (**bis-MTX)** ligand.[198] The size of the protein nanorings can be controlled by varying the length of the linker between the two DHFR units.[199] Taking advantage of this observation, we have been able to construct bivalent (13 amino acid

linker, **13-DHFR^{2}antiCD3**) chemically self-assembled antibody nanorings (CSANs) that mimic the binding and internalization behavior of the parental mAb and octameric (1 amino acid linker, **1-DHFR^{2}antiCD3**) CSANs with superior binding properties to $CD3^+$ HPB-MLT T-leukemia cells compared to the parental mAb.[200] Further, to enhance the utility of this platform we have synthesized a novel bis-MTX dimerizer with an amine containing third arm. Using nucleic acid conjugates to this **amino-bis-MTX** trilinker, we demonstrated the targeted delivery of single and double stranded oligonucleotides to T-leukemia cells.[201] To expand the organizational and delivery capabilities of the CSANS approach beyond oligonucleotides to proteins, we have chosen to explore the utility of protein farnesyltransferase (PFTase) bioorthogonal protein labeling. Previously, our group and a number of other groups have exploited the high specificity of PFTase to site-specifically modify proteins.[7,82,111] PFTase catalyzes the transfer of an isoprenoid group from farnesyl diphosphate (FPP, Figure 5.2A) to a cysteineyl sulfur atom present in a tetrapeptide sequence (denoted as a CaaX-box) positioned at the C-terminus of a protein (Figure 5.2B). Importantly, CaaX-box sequences such as CVIA can be appended to the C-termini of many proteins rendering them efficient substrates for PFTase. Since PFTase can tolerate many simple modifications to the isoprenoid substrate,[40,83,108,202] it can be used to introduce chemoselectively a variety of functional groups into proteins. Previously, we have showed that aldehyde-containing FPP analogs and alkyne-containing FPP analogs can be successfully incorporated into proteins using this strategy.[40,80,82] Consequently, we envisioned that the coupling of a protein modified with an alkyne-containing FPP analog that could be coupled to an analogue of **amino-bis-MTX**, in

which the amino group has been substituted with an azido group, would facilitate incorporation of functional proteins into antibody-nanorings by self-assembly. To achieve this objective, a bifunctional FPP analogue linker was designed with orthogonal reactivity. This linker (**1**) was comprised of an aldehyde group and an alkyne moiety suitable for protein modification (Figure 5.2A). The orthogonal functionalities in the linker could then be used to selectively assemble the antibody-protein nanostructure as illustrated in Figure 5.1.

Experimental Section

Enzymatic studies of FPP-analogue 1 using a continuous fluorescence assay. Enzymatic reaction mixtures contained Tris·HCl (50 mM, pH 7.5), MgCl2 (10 mM), ZnCl2 (10 μM), DTT (5.0 mM), 2.4 μM N-dansyl-GCVIA (2), 0.040 % (w/v) n-dodecyl-ß-D-maltoside, PFTase (80 nM), and varying concentrations of **1** (0-50 μM), in a final volume of 250 μL. The reaction mixtures were equilibrated at 30 °C for 1 min, initiated by the addition of PFTase, and monitored for an increase in fluorescence (λ_{ex}=340 nm, λ_{em}=505 nm) for approximately 20 min. The initial rates of formation of products were obtained as slopes in IU/min using least squares analysis. Corrections were applied to all the rate calculations based on the difference between the fluorescence intensity of the prenylated product and the starting peptide. Assuming 100% conversion, the difference corresponds only to the fluorescence of the total amount of the product. The slope was then divided by the fluorescence difference followed by multiplying by the total concentration of peptide (2.4 μM) which then gives the rate of formation of product in μM/s. It should be noted that the K_M values reported here are actually apparent K_M

values, since the measurements were performed in only a single peptide concentration. The data were fit to a Michaelis-Menten model ($V = \frac{[E_\circ]k_{cat}*[S]}{K_M+[S]}$) using a nonlinear regression program, to determine k_{cat} and K_M.

Enzymatic incorporation of compound 1 into GFP-CVIA (3).[80] Enzymatic reaction mixtures (10 mL) contained Tris·HCl (50 mM, pH 7.5), $MgCl_2$ (10 mM), KCl (30 mM), $ZnCl_2$ (10 μM), DTT (5.0 mM), 7 (2.4 μM), compound **1** (30-50 μM), and PFTase (80-200 nM). After incubation at 30°C for 4 h, the reaction mixture was concentrated using an Amicon Centriprep centrifugation device (10,000 MW cut-off). Next, excess of **1** was removed through a NAP-5 (Amersham) column using Tris·HCl (50 mM, pH 7.5) as the eluant. The subsequent protein concentration was calculated by UV absorbance at 488 nm (ε=55,000 $M^{-1}\cdot cm^{-1}$).

Simultaneous coupling reactions between bifunctional-GFP (4) with azide-bisMTX (6) and aminooxy-TAMRA (5). Aminooxy-TAMRA (**5**) (7 μL of 2.2 mM solution in DMSO) and azide-bisMTX (**6**) (7 μL of 10 mM solution in DMSO) were added to 100 μL **4** (stock solution of 40 μM in PB). $CuSO_4$ (1 mM), TCEP (1 mM), TBTA (100 M) and m-phenylenediamine (40 mM) were added to the mixture and the reaction was allowed to proceed for 15 h. LC-MS analysis of the sample showed >90% conversion of the click reaction and >99% of oxime ligation on the bifunctionalized-protein **4**. The mixture was then purified by a NAP-5 column to remove the excess of reagents.

Evaluation of the chemically induced self-assembly of TAMRA-GFP-bisMTX (7) with dimeric DHFR proteins by HP-SEC. TAMRA-GFP-bisMTX (7) (final

concentration of 10 uM) and dimeric DHFR proteins (final concentration of 10 uM) were mixed and incubated at rt for 2 h. Next the samples were analyzed by HP-SEC.

Confocal microscopy. 0.5 x 10^6 HPB-MLT cells were treated with **7** (1 μM) for control experiments and with self-assembled nanostructures (**TG-13DDCD3** or **TG-1DDCD3**) at either 4 or 37 °C for 1 h in RPMI media. Cells were then pelleted by centrifugation (400 x g, 5 min). After being washed twice with PBS (phosphate buffer saline), cells were incubated on Poly-Prep slides coated with poly-L-Lysine (Sigma) at 4 or 37 °C for 30 min. Cells were then fixed with 4% paraformaldehyde solution for 10 min and washed thrice with PBS. Finally, cells were treated with ProLong Gold Antifade reagent with DAPI (Invitrogen) and a cover slip was applied. After overnight incubation they were imaged by fluorescence confocal microscopy using an Olympus FluoView 1000 BX2 upright confocal microscope.

Flow Cytometry. 1 x 10^6 HPB-MLT cells were treated with either **TG-13DDCD3** or **TG-1DDCD3** (1/0.5/0.1 μM) at 4 °C for 1 h in PBS buffer (containing 0.05 % BSA and 0.1 % sodium azide). Cells were pelleted (400 x g, 10 min), washed twice, resuspended in the supplemented PBS and their fluorescence was analyzed with a FACS Calibur flow cytometer (BD Biosciences). For the positive control experiment, 1 x 10^6 HPB-MLT cells were incubated with 40 nM UCHT-1 (anti-CD3 monoclonal antibody). After 2 h of incubation, cells were washed, resuspended and analyzed by flow cytometry. For cell specific experiments, 1 x 10^6 $CD3^-$ Daudi B lymphoma cells were incubated with the constructs.

Results and Discussion

Compound **1** was synthesized from commercially available compounds, geraniol and 3,5-dihydroxybenzaldehyde in nine steps (Scheme 5.S1). In brief, THP-protected geraniol was initially oxidized at C-8 to a terminal alcohol, followed by bromination of the alcohol group using CBr_4 and PPh_3. The 3,5-alkyne-aldehyde functionalized phenol was made from 3,5-dihydroxybenzaldehyde via Sonogashira Pd–Cu catalyzed cross-coupling reaction[203] in three steps. The modified phenol was attached to the bromide via a substitution reaction using K_2CO_3 as the base. The THP group was removed and the alcohol was converted to the corresponding diphosphate via a direct phosphorylation strategy employing $(HNEt_3)_2HPO_4$ and CCl_3CN as the activating reagent. Subsequent purification by RP-HPLC produced the desired bifunctional aldehyde-alkyne analogue **1** whose structure was confirmed by ^{1}H-NMR, ^{31}P-NMR, and HR-ESI-MS.

Initially, prenylation reactions containing a model peptide, *N*-dansyl-GCVIA (**2**), analog **1**, and PFTase were performed and the reaction was monitored via a continuous fluorescence-based enzyme assay, as previously reported,[159] which demonstrated that compound **1** is in fact an alternative substrate for the enzyme. Next, a kinetic analysis of that reaction was performed using varying concentrations of **1** and *N*-dansyl-GCVIA (**2**) in the presence of PFTase showing that the enzymatic process obeyed saturation kinetics. Steady-state kinetic parameters for the prenylation reaction using the bifunctional aldehyde-alkyne analogue are provided in the Supporting Information section (Figure 5.S3). This analysis revealed that the catalytic efficiency for alternative substrate **1** is reduced relative to that of FPP, with a k_{cat}/K_M value of 0.02 (relative to FPP). We found

that decrease in k_{cat} constituted the major reason for the diminished catalytic efficiency of the analogue, with the k_{cat} value for **1** (k_{cat}=0.0123 s^{-1}) observed to be 42-fold lower than the k_{cat} for FPP (k_{cat}=0.52 s^{-1}), while no substantial difference was observed between the K_M values for the analogue (K_M=2.52 μM) and that of FPP (K_M= 1.71 μM).

With the ability of analogue **1** to be incorporated by PFTase demonstrated, we next evaluated the utility of the analogue for selective protein modification. Accordingly, **1** was incubated with GFP-CVIA (**3**) in the presence of PFTase (5 h, rt). That choice of reaction time was based on our earlier work with the peptide substrate *N*-dansyl-GCVIA (**2**) where a 2 h reaction resulted in complete conversion. Concentration by ultracentrifugation followed by size-exclusion chromatography to remove excess **1** yielded pure bifunctionalized GFP **4** (Figure 5.S5). Completion of the reaction was confirmed by LC-ESI/MS (Figure 5.2C and 5.2D) with the major peak corresponding to prenylated GFP (**4)** and a minor peak corresponding to negligible amount of unreacted GFP-CVIA (**3**) (Figure 5.S5).

In a previous study, we had shown that aldehyde-GFP and alkyne-GFP modified proteins could be derivatized to produce oxime-linked or clicked products, respectively.[40,80] In this study we explored simultaneous oxime and click reactions on a single prenylated-protein. First, to demonstrate the orthogonality of the two reactions, we separately carried out either an oxime ligation or a click reaction on the bifunctionalized protein. Hence, oxime ligation on **4** was performed with an aminooxy-dansyl compound (**S10**) using m-phenylenediamine (mPDA) as the catalyst.[204] LC-MS analysis of the crude ligation reaction mixture confirmed complete conversion to the oxime-linked product in

less than 1 h (Figure 5.S6). The click reaction was tested on **4** with TAMRA-azide (**S11**) using TCEP, TBTA and copper as the catalyst. LC-MS analysis revealed that >80% conversion was obtained within 12 h of reaction, using equimolar concentrations of prenylated-protein and the coupling reagent (TAMRA-azide) as used in the above oxime ligation reaction. Typically the oxime ligation reaction proved to be modestly more efficient than the click reaction (Figure 5.S8). When the oxime and click reactions were carried out simultaneously on the modified protein using equimolar concentrations of reagents (*vide supra*), quantitative conversion for the oxime reaction and ~90% conversion for the click reaction after 15 h (Figure 5.S8) was observed.

After optimizing the conditions for the oxime and click reactions on the prenylated protein **4**, we next prepared **bis-MTX-GFP-TAMRA, 7**. The previously reported **amino-bis-MTX** trilinker was coupled to 5-azidopentanoic acid using standard peptide chemistry to obtain **azido-bis-MTX**, **6**. Next, prenylated protein **4** was incubated with **aminooxy-TAMRA**, **5**, and **azido-bis-MTX**, **6**, for 15 h using the optimized reaction conditions (*vide supra*). Subsequent LC-MS analysis of the reaction mixture revealed the successful simultaneous conjugation of the protein with both **aminooxy-TAMRA** and **azido-bis-MTX** (Figure 5.2E). Overall, the conjugation reactions appear to be highly efficient, since little free starting material was observed by LC-MS. **Bis-MTX-GFP-TAMRA**, **7**, was further purified by HP-SEC.

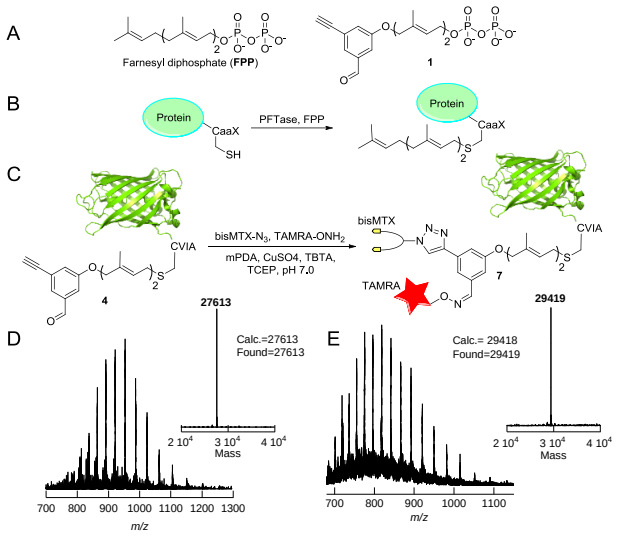


Figure 5.2. A) Structures of FPP and compound **1**. B) Schematic representation of farnesylation reaction with a protein containing a CaaX box at its C-terminus. C) Schematic representation of simultaneous oxime ligation and click reaction of the bifunctionalized prenylated protein **4** with **azido-bis-MTX 6** and **aminooxy-TAMRA 5**. D and E) ESI MS analysis of **4** and **7** with the deconvoluted mass spectra shown in the insets, respectively.

With the bis-MTX- and TAMRA-functionalized GFP-protein (**7**) in hand, we then evaluated the chemically induced self-assembly of this protein with dimeric DHFR proteins by HP-SEC. After simple mixing of the two entities and incubation at rt for 2 h, the samples were analyzed by HP-SEC. The elution profiles of the HP-SEC traces indicated the predominant formation of the internal monomeric species with **13-DHFR^2** and much higher order species with **1-DHFR^2** (Figure 5.S10-11). Having confirmed the self-assembly with non-targeting DHFR proteins, we then moved to construct the **TAMRA-GFP-DHFR^{2}antiCD3** (**TG-DDCD3**) protein nanostructures by mixing **13-DHFR^{2}antiCD3** (**13DDCD3**) and **1-DHFR^{2}antiCD3** (**1DDCD3**) with **7**. The targeting proteins generated a similar pattern in the SEC profile (Figure 5.3).

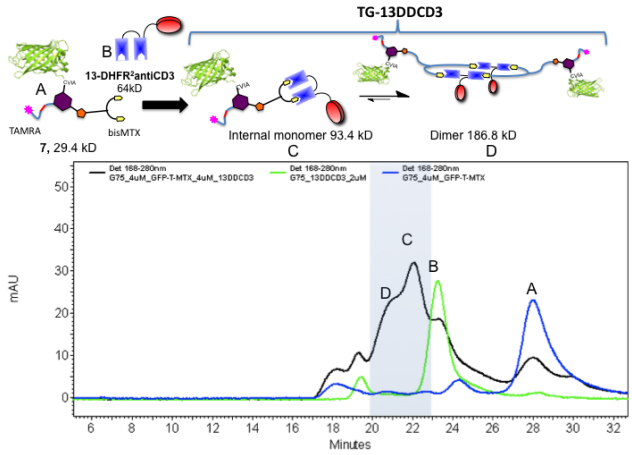


Figure 5.3. Self-assembly of antibody nanorings observed by SEC (**13-DHFR^{2}antiCD3** with **7**). Blue curve: bis-MTX-GFP-TAMRA (**7**, Peak **A**). Green curve: monomeric **13-DHFR^{2}antiCD3** (Peak **B**). Black curve: induced oligomerization of **13-DHFR^{2}antiCD3** with **7** indicating the major products as the internal monomer (Peak **C**) and dimer (Peak **D**); Shaded area indicates the collected peak for internalization studies.

As expected, TAMRA-GFP-DHFR CSANs prepared with **13-DHFR^{2}antiCD3**, exhibited predominantly internal dimer, dimer and small amount of higher order species. The peaks eluting between 20 –22.5 min (**TG-13DDCD3**) were collected for cellular internalization studies (Figure 5.3). Similar experiments were performed with **1-DHFR^{2}antiCD3** as well (Figure 5.S13 and 5.S14).

In order to confirm the functionality of the TAMRA-GFP anti-CD3 CSANs (**TG-DDCD3**'s), we determined their ability to specifically deliver the protein nanostructure to $CD3^+$ T-Leukemia cells. The cells were treated with **TG-13DDCD3** and incubated for 1 h and 24 h (at 4 °C and 37 °C) and evaluated by confocal laser scanning microscopy for cellular internalization. Both the GFP and TAMRA chromophores were excited

individually to observe the GFP and TAMRA emission. Next, GFP was excited and TAMRA emission was observed by FRET between GFP and TAMRA.

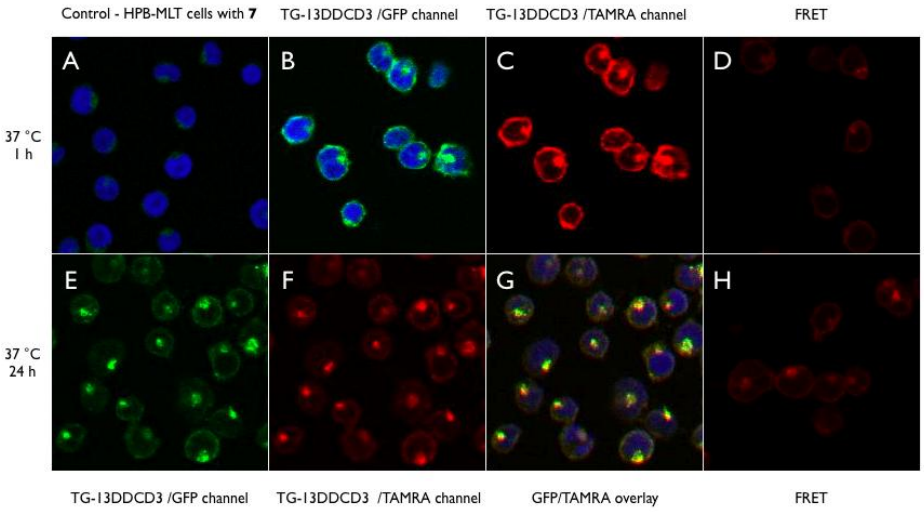


Figure 5.4. A) Internalization studies of **TG-13DDCD3** with HPB-MLT T-leukemia cells at 37 °C. (A) HPB-MLT cells treated with bis-MTX-GFP-TAMRA (**7**) for 1 h (control, blue=DAPI nuclear stain); (B-D) **TG-13DDCD3** incubated with $CD3^+$ T-leukemia cells for 1 h; (E-H) **TG-13DDCD3** incubated with $CD3^+$ T-leukemia cells for 24 h. (D&H) FRET between GFP and TAMRA was observed by exciting the cells with 488 nm laser and a BA560/620 nm emission filter.

Confocal microscopy images confirmed the delivery of the nanostructures (**TG-13DDCD3** and **TG-1DDCD3)** into the cells via an energy dependent endocytosis process (Figure 5.4 and Figures S12-14). At 4 °C, both GFP and TAMRA fluorescence was observed on the cell membranes, (Figure 5.S12) and at 37 °C distinct punctate structures (Figure 5.4B&C) were observed within the cells indicating the energy dependent uptake of both GFP and TAMRA. Overlay of the GFP/TAMRA channels indicated that the two fluorophores resided in the same location, while the positive FRET signal was consistent with the GFP/TAMRA conjugate, **7**, having remained intact in the cells during the time

period of the experiment (24 h) (Figure 5.4E-H). These results are consistent with our previous observations with antiCD3 scFv CSAN's.[200] As was noted in earlier studies with antiCD3 CSAN's, flow cytometric analysis (Figures 5.S15,16), revealed that the nanostructure bind specifically to $CD3^+$ HPB-MLT cells but not $CD3^-$ Daudi cells (Figure 5.S19). When compared, little difference by flow cytometry was observed between the binding by CSANs prepared from **TG-13DDCD3** and **TG-1DDCD3** to $CD3^+$ HPB-MLT cells, despite the supposed higher valency of **TG-1DDCD3** based CSANs (Figures S17-18). This discrepancy may result from a lack of significant differences in the size of CSANs prepared from either monomer. The observation that the hydrodynamic radii of CSANs prepared from **TG-13DDCD3** and **TG-1DDCD3** are similar supports this hypothesis.

Conclusions

In summary, we have demonstrated that enzymatic protein modification and bis-MTX driven chemical dimerization can be used to prepare complex multi-protein assemblies. Using these tools, we have prepared a defined assembly composed of three proteins; a bioorthogonally chemically modified protein and a fusion protein. As a model system, we chose to prepare CSANs by chemical dimerization from bis-MTX conjugated GFP and anti-CD3 single chain antibodies using a DHFR-DHFR fusion protein as a scaffold for a targeting protein (13DDCD3 and 1DDCD3). Confocal microscopy and Flow cytometry experiments confirmed that the CSANs are fully capable of delivering labeled-GFP specifically to $CD3^+$ cells and that the GFP-TAMRA conjugate remains intact once internalized. The multifaceted chemical approach reported here for protein assembly sets

the stage for future studies that will examine the ability of our self-assembly approach to be applied to the development of novel multivalent immunotoxins, as well as the construction of bispecific antibodies, in addition to the simultaneous monitoring of their *in vivo* activity by fluorescence or radiochemical methods.

Funding Sources

Financial support for these studies through NIH Grants GM058842 (M.D.D), GM084152 (M.D.D), CA120116 (C.R.W.) and the University of Minnesota is gratefully acknowledged**.**

Supporting Information

General. All synthetic reactions were carried out at rt and stirred magnetically unless otherwise noted. TLC was performed on precoated (250 mm) silica gel 60 F-254 plates (Merck). Plates were visualized by staining with $KMnO_4$ or with a hand-held UV lamp. Flash chromatography was performed using a Biotage® instrument. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. ^1H NMR spectra were obtained at 500 MHz; ^{13}C NMR spectra were obtained at 125 MHz. All NMR spectra were acquired on Varian instruments at 25°C. Chemical shifts are reported in ppm and *J* values are in Hz. Fluorescence assay data were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer. MS spectra for synthetic reactions were obtained on a Bruker BioTOF II instrument. Yeast PFTase was prepared as previously described.[173] Protein LC/MS analyses were performed using a Waters Synapt G2 Quadropole TOF mass spectrometer instrument. MALDI-MS analyses were performed with a Bruker MALDI TOF spectrometer Instrument. Preparative HPLC separations were performed by using a Beckman model 127/166 instrument, equipped with a UV detector and a Phenomenex C_{18} column (Luna, 10 µm, 10x250 mm). Vydac 218TP1010 column was used for preparative RP-HPLC. Size Exclusion Chromatography (SEC) was performed on a System Gold 126/168 HPLC system (Beckman-Coulter, Fullerton, CA, USA) connected to a Superdex G75 or G200 column (GE Healthcare Life sciences) with P500 buffer (0.5 M NaCl, 50 mM KH_2PO_4, 1 mM EDTA, pH 7) as the mobile phase (flow rate 0.5 mL/min). All solvents were of HPLC grade. All other reagents were from Sigma Aldrich.

Abbreviations.

GFP, green fluorescent protein;

PFTase, protein farnesyl transferase;

Bis-MTX, bis-methotrexate;

mPDA, m-phenylenediamine;

FPP, Farnesyl diphosphate;

DTT, dithiothreitol;

ESI-MS, electrospray ionization mass spectrometry;

RP-HPLC, reversed-phase high-pressure liquid chromatography;

PB, phosphate buffer;

DMF, dimethylformamide;

TEA, triethylamine;

Tris, tris(hydroxymethyl)aminomethane;

EDTA, Ethylenediaminetetraacetic acid;

SEC, size exclusion chromatography

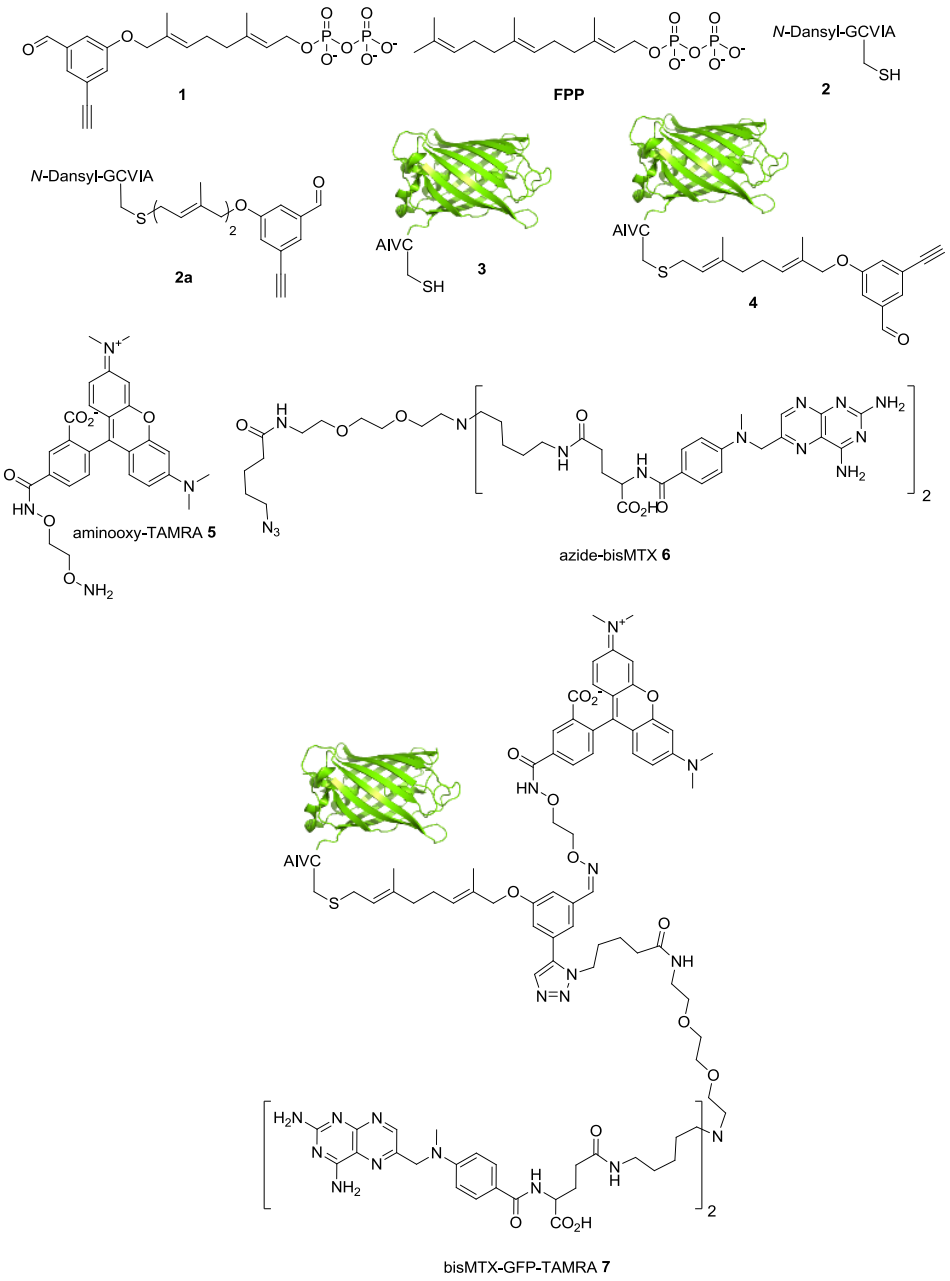


Figure 5.S1. Structures of compounds 1-7.

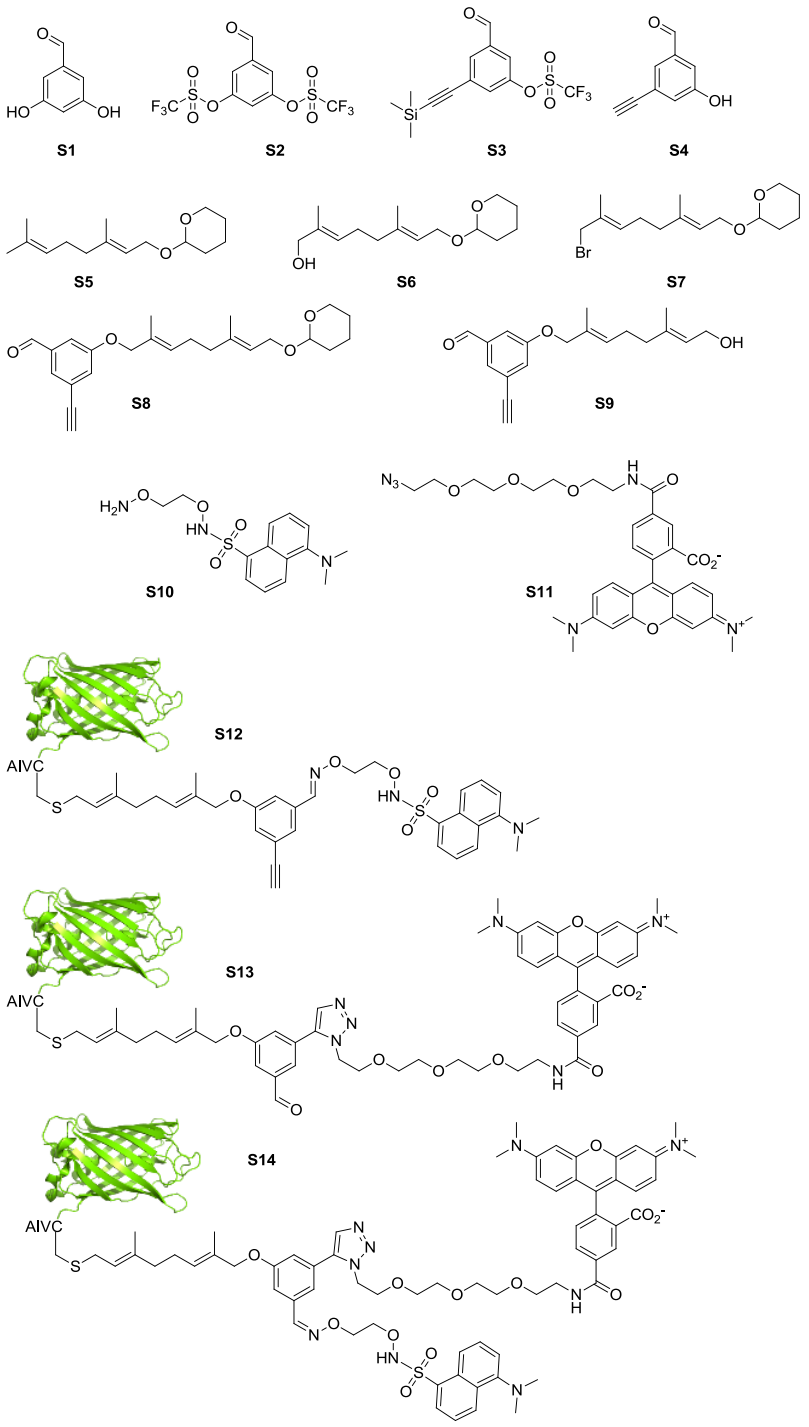


Figure 5.S2. Structures of compounds S1-S14.

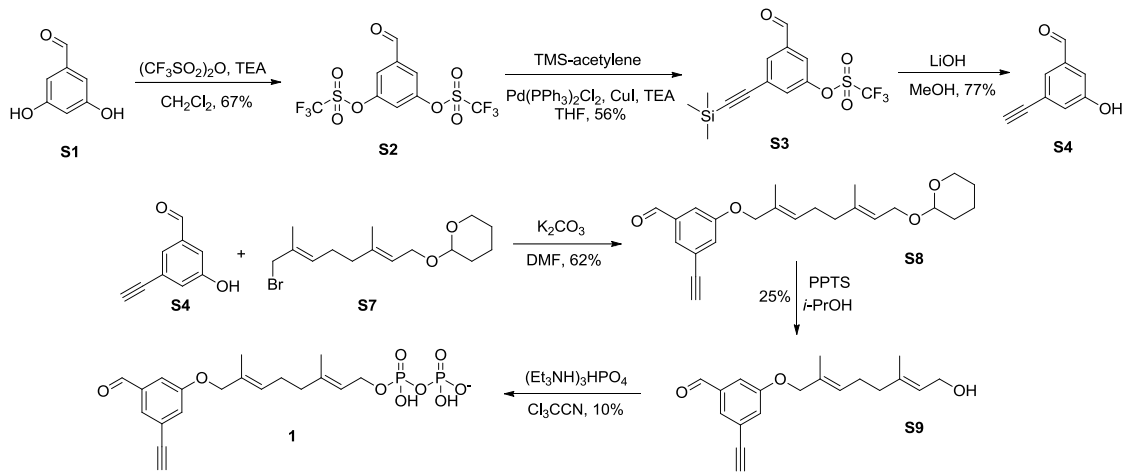
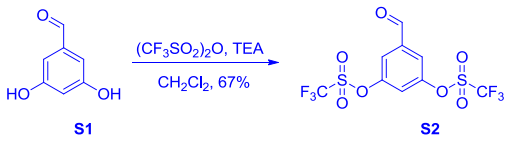


Figure 5.S3. Schematic synthesis of compound **1**.

Synthesis of compound 1.

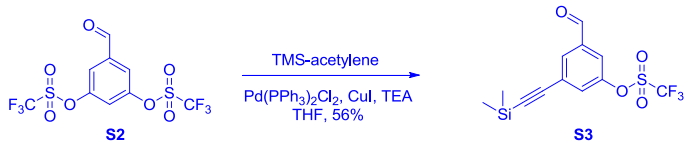
5-formyl-1,3-phenylene bis(trifluoromethanesulfonate) (S2).



Compound **S2** was synthesized according to modified literature procedures.[155,205] Under N_2, 3,5-dihydroxy-benzaldehyde (2.00 g, 14.5 mmol, 1.0 eq) was dissolved in CH_2Cl_2 (20 mL) and TEA (43.4 mmol, 6.05 mL, 4.4 g). The mixture was cooled in an ice bath at 0 °C and a solution of $(CF_3SO_2)_2O$ (8.9 g, 31.5 mmol, 2.2 eq) in CH_2Cl_2 (10 mL) was added dropwise. The reaction mixture was stirred for another 2 h and allowed to warm up to rt during this period. H_2O (50 mL) was added and the product was extracted with CH_2Cl_2 (2×30 mL). The combined organic layers were washed with 1 M HCl (30 mL), H_2O (30 mL) and brine (20 mL), dried over Na_2SO_4 and then solvent was removed *in vacuo*. The brown residue was purified by silica gel flash column chromatography (Hex/EtOAc 4:1) to yield **S2** (3.86 g, 67%) as a slightly yellowish powder. ^{1}H NMR (δ,

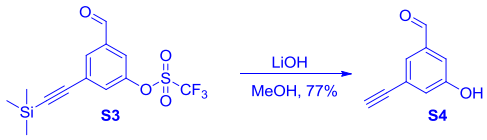
$CDCl_3$): 10.021 (s, 1H), 7.845 (d, J = 2.0 Hz, 2H), 7.484 (t, J = 2.2 Hz, 1H). ^{13}C NMR (δ, $CDCl_3$): 187.527, 158.932, 150.049, 121.983, 120.774, 117.337.

3-formyl-5-((trimethylsilyl)ethynyl)phenyl trifluoromethanesulfonate (S3).



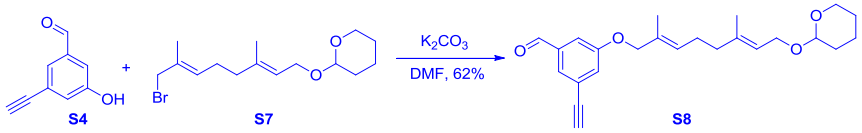
Compound **S3** was synthesized according to modified literature procedures.[155,205] Compound **S2** (3.6 g, 8.95 mmol, 1.0 eq) was dissolved in THF (40 mL) and TEA (20 mL). $Pd(PPh_3)_2Cl_2$ (63.5 mg, 0.09 mmol, 0.01 eq) and CuI (34.7 mg, 0.18 mmol, 0.02 eq) were added to the reaction mixture. The mixture was cooled in an ice bath at 0°C and a solution of TMS-acetylene (1.27 mL, 8.95 mmol, 1.0 eq) in THF (20 mL) was slowly added dropwise. The reaction was monitored by TLC. The mixture was stirred for another 2 h during which time the temperature was allowed to rise to rt. The solvents were removed *in vacuo* and sat. NH_4Cl (30 mL) and EtOAc (50 mL) were added. The layers were partitioned and the aqueous layer was extracted with EtOAc (50 mL). The combined organic layers were washed with H_2O (50 mL) and brine (50 mL), dried over Na_2SO_4 and solvent was removed *in vacuo.* Purification by silica gel flash column chromatography (Hex/EtOAc 30:1) gave **S3** (1.72 g, 4.9 mmol, 55.5%) as a colorless oil. ^1H NMR (δ, $CDCl_3$): 9.965 (s, 1H), 7.948 (d, J = 1.0 Hz, 1H), 7.692 (dd, J = 1.0, 1.0 Hz, 1H), 7.564 (dd, J = 1.0, 1.5 Hz, 1H), 0.254 (s, 9H). ^{13}C NMR (δ, $CDCl_3$): 189.185, 149.601, 138.101, 133.285, 129.713, 120.926, 100.941, 99.540, 79.803, 68.189, -0.383.

3-ethynyl-5-hydroxybenzaldehyde (S4).



Compound **S4** was synthesized according to modified literature procedures.[155,205] **S3** (1.7 g, 4.85 mmol, 1.0 eq) was dissolved in MeOH (15 mL). A solution of LiOH (1.0 g, 23.8 mmol, 5 eq) in MeOH (15 mL) was added at 0°C, and next the temperature was allowed to rise to rt overnight. The mixture was extracted with Et_2O (3 × 15 mL) and the combined organic layers were washed with H_2O (10 mL) and brine (10 mL) and dried over Na_2SO_4. The solvent was removed *in vacuo* and purification by silica gel flash column chromatography (Hex/Et_2O 2:1) gave **S4** (0.55 g, 77%) as a white powder. ^{1}H NMR (δ, $CDCl_3$): 9.895 (s, 1H), 7.537 (dd, J = 1.5, 1.5 Hz, 1H), 7.336 (dd, J = 1.0, 1.0 Hz, 1H), 7.213 (dd, J = 1.0, 1.0 Hz, 1H), 6.094 (s, 1H), 3.126 (s, 1H); ^{13}C NMR (δ, $CDCl_3$): 191.476, 156.345, 137.802, 127.061, 124.955, 115.343, 113.589, 81.798, 78.762.

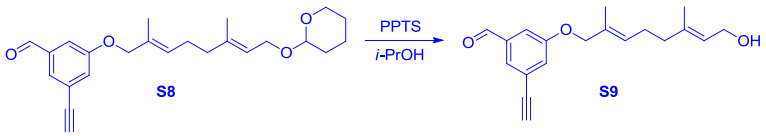
3-(((2E,6E)-2,6-dimethyl-8-((tetrahydro-2H-pyran-2-yl)oxy)octa-2,6-dien-1-yl)oxy)-5-ethynylbenzaldehyde (S8).



Compounds **S5**, **S6** and **S7** were prepared as previously described.[40,206] Bromide **S7** (0.036 g, 1.136 mmol) and **S3** (0.292 g, 2.0 mmol) were dissolved in DMF (15 mL) in a

flame dried 50 mL flask. K_2CO_3 (0.50 g, 3.62 mmol) was added to the reaction flask, then left to stir at 100 °C for 3 h until TLC analysis (2:1 Hex:EtOAc v/v) indicated almost complete conversion to the product. The solvent was removed *in vacuo* and the crude product was further purified by silica gel column chromatography with gradient elution (Hex:Et_2O) from 1:0 (v/v) going to 5:1 (v/v) to afford 0.27 g of compound **S8** (0.7 mmol, 62% yield) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.910 (s, 1H), 7.542 (dd, J = 1.0 Hz, 1H), 7.373 (dd, J = 1.0 Hz, 1H), 7.265 (dd, J = 1.0 Hz, 1H), 5.538 (t, J = 7.0 Hz, 1H), 5.357 (t, J = 6.5 Hz, 1H), 4.610 (m, 1H), 4.420 (s, 2H), 4.235 (dd, J = 9.5, 6.5 Hz, 2H), 4.004 (m, 1H), 3.95–3.83 (m, 1H), 3.495 (m, 1H), 3.135 (s, 1H) 2.201 (m, 2H), 2.081 (t, J = 7.5 Hz, 2H), 1.714 (s, 3H), 1.671 (s, 3H), 1.66–1.46 (m, 5H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 191.132, 159.165, 139.845, 139.436, 137.626, 130.257, 129.160, 127.406, 126.765, 124.698, 120.982, 114.326, 97.817, 82.006, 78.458, 63.552, 62.238, 38.865, 30.639, 25.905, 25.417, 19.552, 16.350, 13.779. HR-ESI-MS calcd for $C_{24}H_{30}O_4Na$ $[M+Na]^+$ 405.2037, found 405.2051.

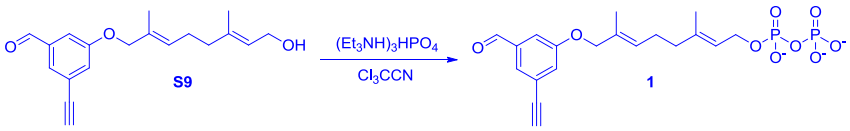
3-ethynyl-5-(((2E,6E)-8-hydroxy-2,6-dimethylocta-2,6-dien-1-yl)oxy)benzaldehyde (S9).



Protected alkyne-aldehyde **S8** (0.25 g, 0.65 mmol) was dissolved in *i*-PrOH (15 mL) in a 25 mL flask. PPTS (10 mg) was added as catalyst. The reaction was then refluxed at 75 °C for 4 h, when TLC analysis indicated complete conversion to the product. It was then quenched by adding sat. $NaHCO_3$ (5 mL) and EtOAc (50 mL). The organic layer was

then separated and dried over Na_2SO_4. The solvent was removed *in vacuo* and afforded 49 mg of compound **S9** (25% yield) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.901 (s, 1H), 7.536 (dd, J = 1.0 Hz, 1H), 7.365 (dd, J = 1.0 Hz, 1H), 7.257 (dd, J = 1.0 Hz, 1H), 5.519 (t, J = 7.0 Hz, 1H), 5.386 (t, J = 6.5 Hz, 1H), 4.419 (s, 2H), 4.126 (d, J = 7.0, 2H), 3.115 (s, 1H) 2.195 (m, 2H), 2.064 (t, J = 7.0 Hz, 2H), 1.701 (s, 3H), 1.657 (s, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 191.476, 156.345, 139.640, 141.820, 137.802, 127.061, 124.955, 124.481, 115.343, 113.573, 81.798, 80.220, 78.762, 68.229, 65.987, 27.067, 22.822, 15.116, 14.676.

(2E,6E)-8-(3-ethynyl-5-formylphenoxy)-3,7-dimethylocta-2,6-dien-1-yl diphosphate (1).



Alcohol **S9** (25 mg, 0.084 mmol, 1 eq) was added to CCl_3CN (50.5 uL, 0.5 mmol, 6 eq) in a 25 mL flask. In a separate 5 mL flask, $(Et_3NH)_2HPO_4$ salt (0.25 mmol, 75.6 mg, 3 eq) was added to dry CH_3CN (3 mL), and placed in an oil bath at 30 °C for 5 min to dissolve the salt. This solution was added drop-wise to the mixture of alcohol **S9** and CCl_3CN solution over 3 h at rt, and was left to stir for an additional 50 min at rt. The slow addition of salt solution to the reaction flask was critical to significantly increase the yield. The solvent was removed *in vacuo* and NH_4HCO_3 (25 mM, 5 mL) was added to the resulting solution and a white precipitate was formed. The solution was filtered and purified by RP-HPLC with a semi-preparative column under the following conditions: detection at 254 nm; flow rate at 5.0 mL·min^{-1}; 5 mL injection loop; solvent A: 25 mM

NH_4HCO_3 in H_2O, solvent B: CH_3CN. compound **1** eluted from 20–25% solvent B. Fractions containing pure **1** were collected and lyophilized. The resulting salt was dissolved in D_2O and its concentration was measured following a previously established NMR-based quantification[5] to yield 4 mL of 2.0 mM solution of **1** (3.84 mg, 10% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.712 (s, 1H), 7.549 (dd, J = 1.0 Hz, 1H), 7.371 (dd, J = 1.0 Hz, 1H), 7.297 (dd, J = 1.0 Hz, 1H), 5.458 (t, J = 7.0 Hz, 1H), 5.265 (t, J = 7.0 Hz, 1H), 4.432 (s, 2H), 4.279 (d, J = 6.5, 2H), 3.436 (s, 1H) 2.077 (m, 2H), 1.947 (t, J = 7.5 Hz, 2H), 1.553 (s, 3H), 1.532 (s, 3H). ^{31}P NMR: (121 MHz, D_2O) δ -6.465 (d, J = 17.2, 1P), -3.505 (d, J = 16.5, 1P). HR-ESI-MS calcd for $C_{15}H_{26}O_8P_2$ $[M-H]^-$ 395.1025, found 395.0907.

Enzymatic studies of FPP-analogue 1 using a continuous fluorescence assay.

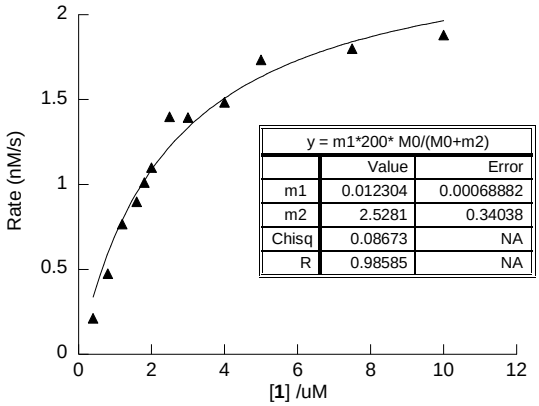


Figure 5.S4. Fluorescence-based PFTase enzyme assay for prenylation of model peptide **3** using varying concentrations of **1**.

GFP-CVIA: Protein was prepared as previously described with one modification.[80,136] In the final phenyl sepharose chromatography step, after the protein was loaded onto the column and washed with buffer as explained in the original work, the protein was eluted from column by adding water instead of buffer.

LC-MS analysis of GFP for determination of prenylation efficiency. Purified prenylated GFP was analyzed by LC-MS to ensure complete prenylation. Proteins were stored in Tris·HCl (50 mM, pH 7.5) prior to injection into the LC-MS instrument. The LC-MS method used was a gradient consisting of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

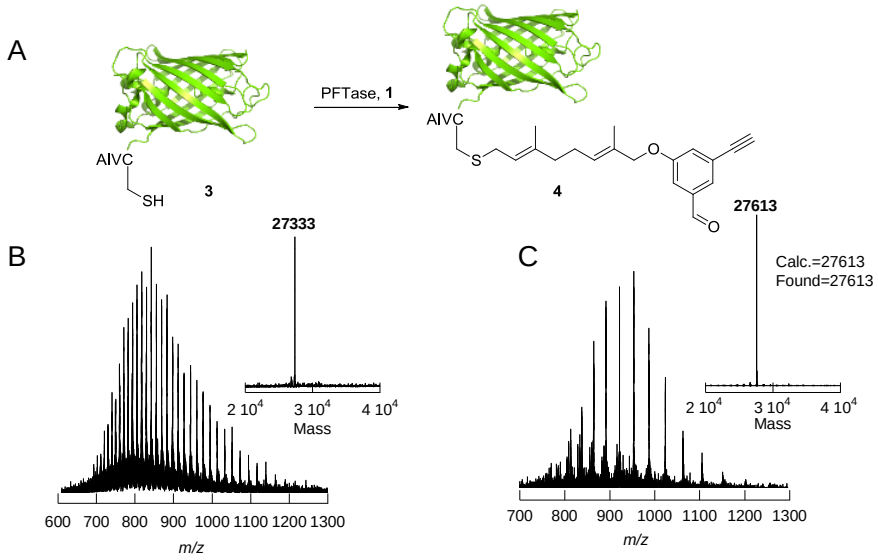


Figure 5.S5. A) Schematic representation of prenylation of GFP-CVIA (**3**) with FPP-analog **1** to yield functionalized GFP **4**. ESI mass spectra of B) GFP-CVIA (**3**) and C) GFP prenylated with bifunctional analogue **1** to yield **4**, showing successful prenylation of GFP **3**. The deconvoluted mass spectra are shown in the insets.

Coupling reaction between bifunctionalized-GFP (4) with aminooxy-dansyl (S10). Aminooxy-dansyl (**S10**) (3.2 µL of 10 mM solution in DMSO) was added to 42 µL of **4**

(stock solution of 60 μM in PB). PB (1 M, pH 7.0, 2.5 μL) was added and the reaction was initiated by adding 40 mM m-phenylendiamine as a catalyst and was allowed to proceed for 2 h at rt. The mixture was then purified by a NAP-5 column to remove excess dye. LS-MS analysis of the sample showed only oxime-ligated protein and no free aldehyde was detected suggesting that the reaction had proceeded to completion.

LC-MS analysis of GFP for determination of oxime ligation efficiency. The crude oxime ligation reaction of bifunctionalized-GFP **4** and aminooxy-dansyl **S10** was analyzed by LC-MS to ensure complete ligation reaction. The LC-MS method used was a gradient of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

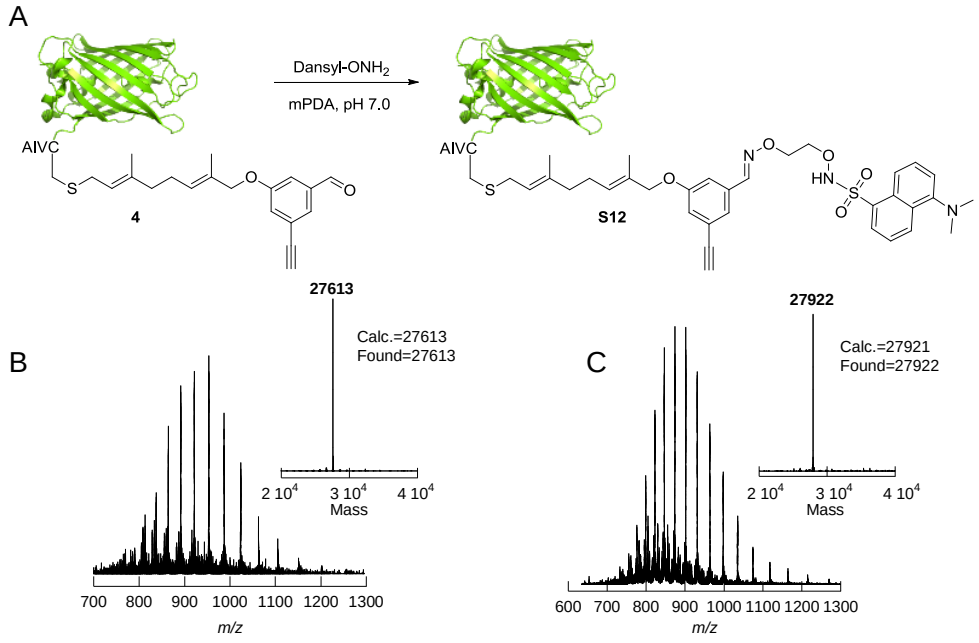


Figure 5.S6. A) Schematic representation of oxime ligation reaction between bifunctionalized-GFP (**4**) and aminooxy-dansyl (**S10**) to yield oxime-GFP (**S12**). ESI mass spectra of B) pure **4** and C) oxime-GFP **4**, indicating successful oxime ligation reaction. The deconvoluted mass spectra are shown in the insets.

Coupling reaction between bifunctionalized-GFP (4) with azide-TAMRA (S11). Azide-TAMRA (**S11**) (7 µL of 2.2 mM solution in DMSO) was added to 100 µL **4** (stock solution of 40 µM in PB). CuSO4 (1 mM), TCEP (1 mM), TBTA (100 M) were added and the reaction was allowed to proceed for 3 h. LC-MS analysis of the sample showed the presence of both the clicked protein **S13** and the free alkyne-protein **4**. The ratio of free alkyne-protein **4** to its respective clicked product **S13**, was ~2, indicating only ~35% completion of click reaction had been achieved within this time and range of reactant concentrations. The reaction was repeated using 50% more azide-TAMRA **S11** (225 µM) and was allowed to proceed for 16 h. LC-MS analysis of the reaction mixture showed >90% conversion of **4** to its respective clicked product **S13**.

LC-MS analysis of GFP for determination of click ligation efficiency. The crude click ligation reaction of bifunctionalized-GFP **4** and azide-TAMRA **S11** was analyzed by LC-MS to evaluate the conversion of the ligation reaction. The LC-MS method used was a gradient consisting of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

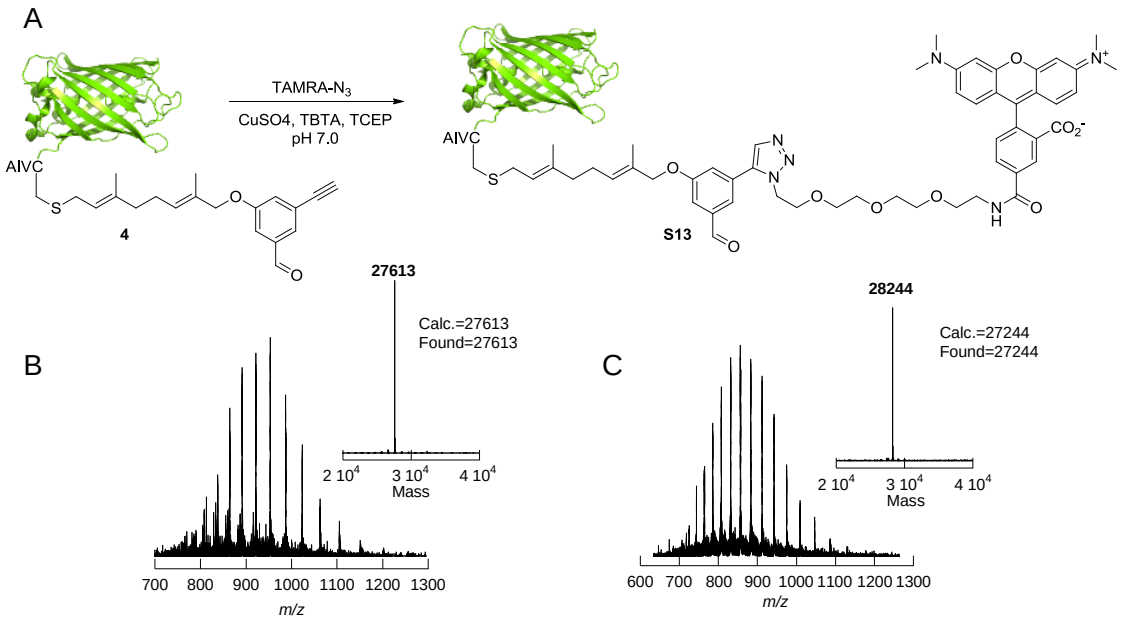


Figure 5.S7. A) Schematic representation of click reaction between the bifunctionalized-GFP **4** with azide-TAMRA **S11** to yield labeled GFP **S13**. ESI mass spectra of B) pure **4** and C) **S13**, showing successful click reaction between **S11** and **4**. The deconvoluted mass spectra are shown in the insets.

Simultaneous coupling reaction between bifunctional-GFP 4 with azide-TAMRA S11 and aminooxy-dansyl S10. Azide-TAMRA (**S11**) (7 µL of 2.2 mM solution in DMSO) and aminooxy-dansyl (**S10**) (7 µL of 10 mM solution in DMSO) were added to 100 µL **4** (stock solution of 40 µM in PB). $CuSO_4$ (1 mM), TCEP (1 mM), TBTA (100 M) and m-phenylenediamine (40 mM) were added and the reaction was allowed to proceed for 15 h. LC-MS analysis of the sample showed >90% conversion of click reaction and >99% of oxime ligation on the protein.

LC-MS analysis of GFP for determination of simultaneous ligations reaction efficiency. Crude reaction mixture of simultaneous click and oxime ligations of bifunctionalized-GFP **4** with azide-TAMRA **S11** and aminooxy-dansyl **S10** was analyzed

by LC-MS to evaluate conversion of ligation reactions. The LC-MS method used was a gradient of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

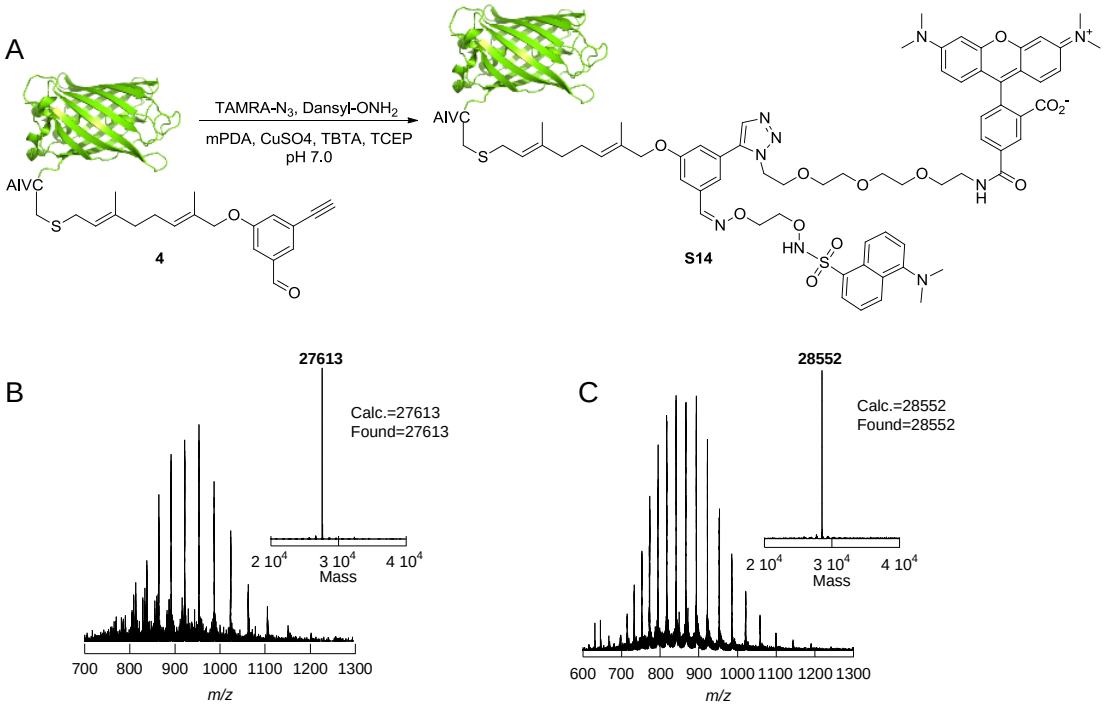
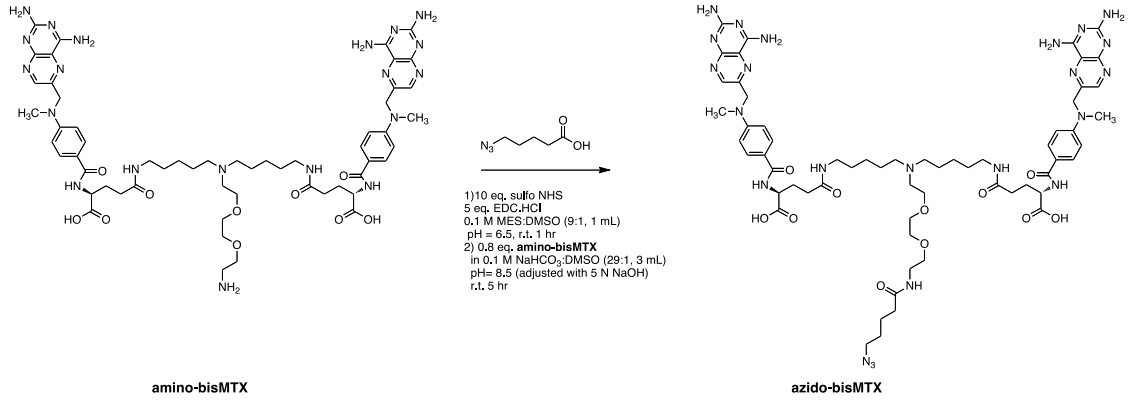


Figure 5.S8. A) Schematic representation of simultaneous click and oxime reactions between the bifunctionalized-GFP (**4**) with aminooxy-dansyl **S10** and azide-TAMRA **S11** to yield GFP **S14**. ESI mass spectra of B) pure **4** and C) **S14**, showing successful simultaneous click and oxime reactions between **S10**, **S11** and **4**. The deconvoluted mass spectra are shown in the insets.

LC-MS analysis of GFP for determination of simultaneous ligations reaction efficiency. Crude reaction mixture of simultaneous click and oxime ligations of bifunctionalized-GFP **4** with aminooxy-TAMRA **5** and azide-bisMTX **6** was analyzed by LC-MS to evaluate conversion of ligation reactions. The LC-MS method used was a

gradient of 0–100% solvent A (H_2O, 0.1% HCO_2H) to B (CH_3CN, 0.1% HCO_2H) in 25 min.

Synthesis of azido-bisMTX.



To a solution of *N*-hydroxysulfosuccinimide (43.4 mg, 200 μmol) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 19.2 mg, 100 μmol) in 0.1 M MES buffer:DMSO (9:1, 1 mL) at pH 6.5 was added 5-azidopropionic acid (2.9 mg, 20 μmol) and the mixture was stirred for 1 h at rt followed by addition of amino-bisMTX[207] (19.1 mg, 16 μmol) in 0.1 M $NaHCO_3$:DMSO (29:1, 3 mL) to the reaction mixture. The reaction mixture was stirred at rt in dark for 5 h. Next, the crude reaction mixture was purified by RP-HPLC using a C8-semi-preparative HPLC column with an H_2O/0.1% TFA and CH_3CN/0.1% TFA gradient. The reation gave 12 mg (83%) of pure product. ESI-MS calcd for $C_{61}H_{85}N_{23}O_{11}$ $[M+3H]^{3+}$ 439.5599, found 439.4167.

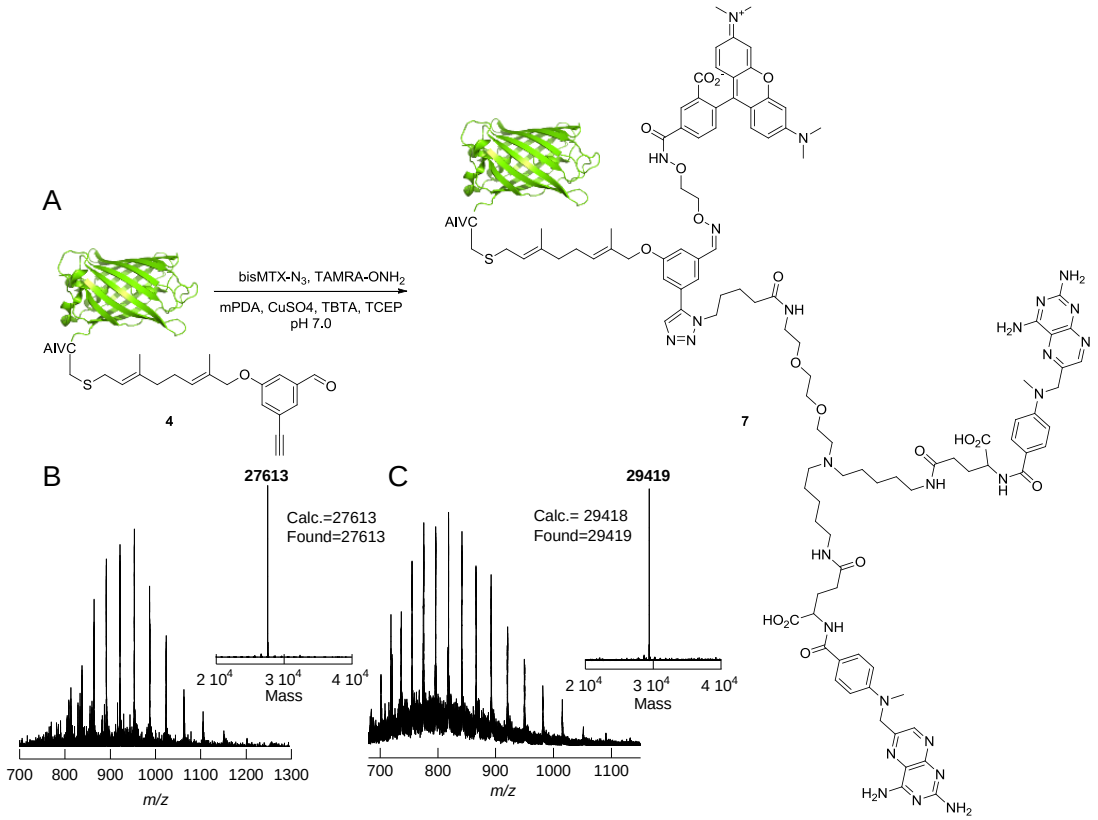


Figure 5.S9. A) Schematic representation of simultaneous click and oxime reactions between the bifunctionalized-GFP **4** with aminooxy-TAMRA **5** and azide-bisMTX **6** to yield GFP **7**. ESI mass spectra of B) pure **4** and C) **7**, showing successful simultaneous click and oxime reactions between **4**, **5** and **6**. The deconvoluted mass spectra are shown in the insets.

Evaluation of the chemically induced self-assembly of TAMRA-GFP-bisMTX (7) with dimeric DHFR proteins by HP-SEC.

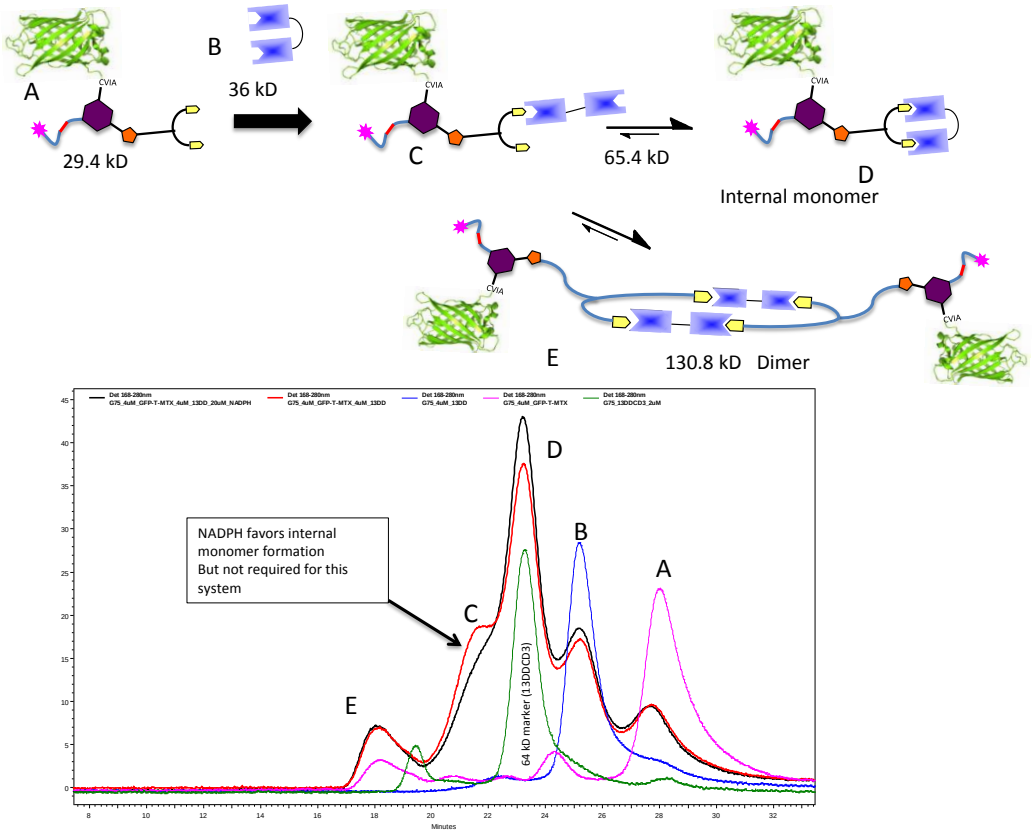


Figure 5.S10. Self-assembly of nanorings observed by SEC (**13-DHFR**^{2} with **7**). Pink curve: bisMTX-GFP-TAMRA (**7,** Peak **A**). Blue curve: monomeric **13-DHFR**^{2} (Peak **B**). Black and Red curves: induced oligomerization of **13-DHFR**^{2} with **7** indicating the major products as the internal monomer (Peak **D**) and dimer (Peak **C**).

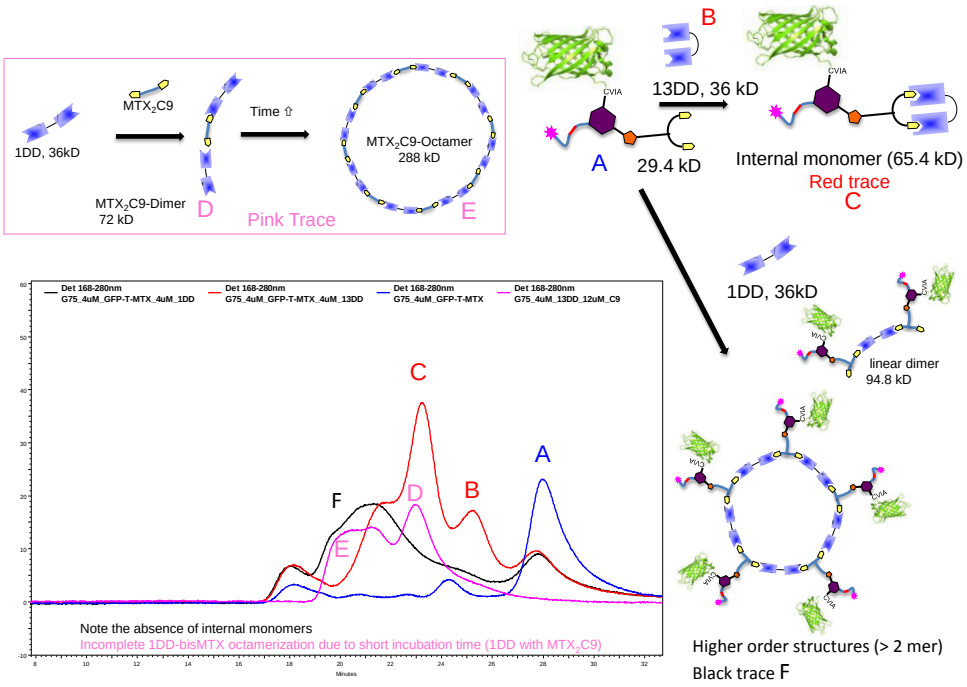


Figure 5.S11. Self-assembly of nanorings observed by SEC (**1-DHFR^2** with **7**). Blue curve: bisMTX-GFP-TAMRA (**7,** Peak **A**). Red curve: induced oligomerization of **13-DHFR^2** with **7** indicating the major products as the internal monomer (Peak **C**). Black curve: induced oligomerization of **1-DHFR^2** with **7** indicating the major products as the much higher order species. Pink curve: oligomerization of **1-DHFR^2** with bisMTX-C9 dimerizer for comparison.

Confocal microscopy.

Laser and filter settings for fluorescence experiments

Laser (nm)	Excite	Ex. Filter	Em. Filter	Color on the image
405	DAPI	DM405/488	BA430/470	Blue
488	GFP	DM405/488/543/635	BA505/525	Green
543	TAMRA	DM405/488/543/635	BA560/660	Red

Laser and filter settings for FRET between GFP and TAMRA

Laser (nm)	Excite	Ex. Filter	Em. Filter	Color on the image
488	GFP	DM405/488	BA560/620	Red

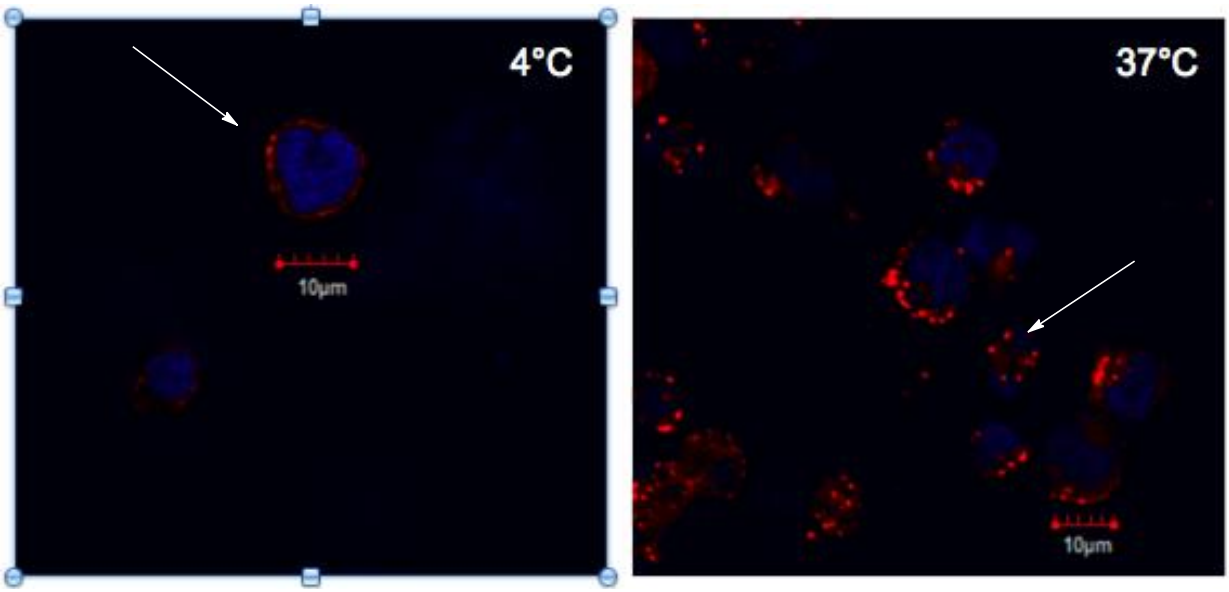


Figure 5.S12. Imaging of HPB-MLT cells treated with **TG-1DDCD3** at either 4 °C (Left) or 37 °C (Right). Cells were treated for 1 h. Arrows indicate red punctuates on the cell membrane at 4 °C (Left) or within the cell at 37 °C (Right) corresponding to the FRET between GFP and TAMRA using above-mentioned filter settings. Magnification: 60x.

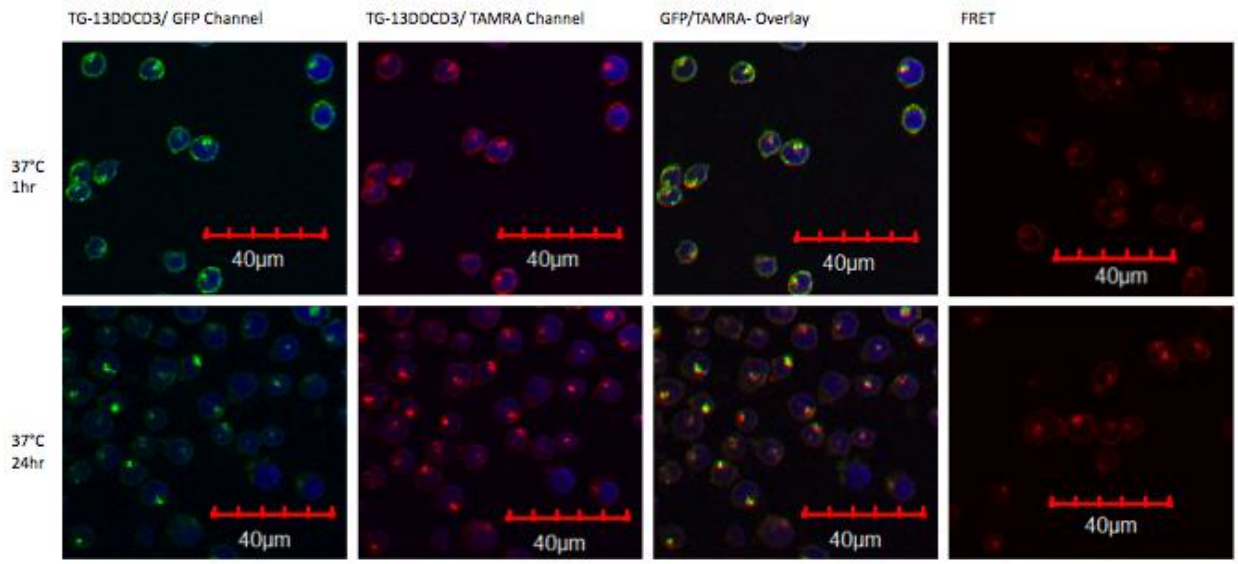


Figure 5.S13. Imaging of HPB-MLT cells treated with **TG-13DDCD3** at 37 °C for 1 h (top panel) and for 24 h (bottom panel). A & E: GFP channel; B & F: TAMRA channel; C &G: Overlay of the GFP and TAMRA channels indicate the intact assembled protein (co-localized yellow punctuates); D & H: Observed FRET between GFP and TAMRA using above-mentioned filter settings. (FRET images were acquired right after GFP and TAMRA imaging and shows a different area of the same slide) Magnification: 40x.

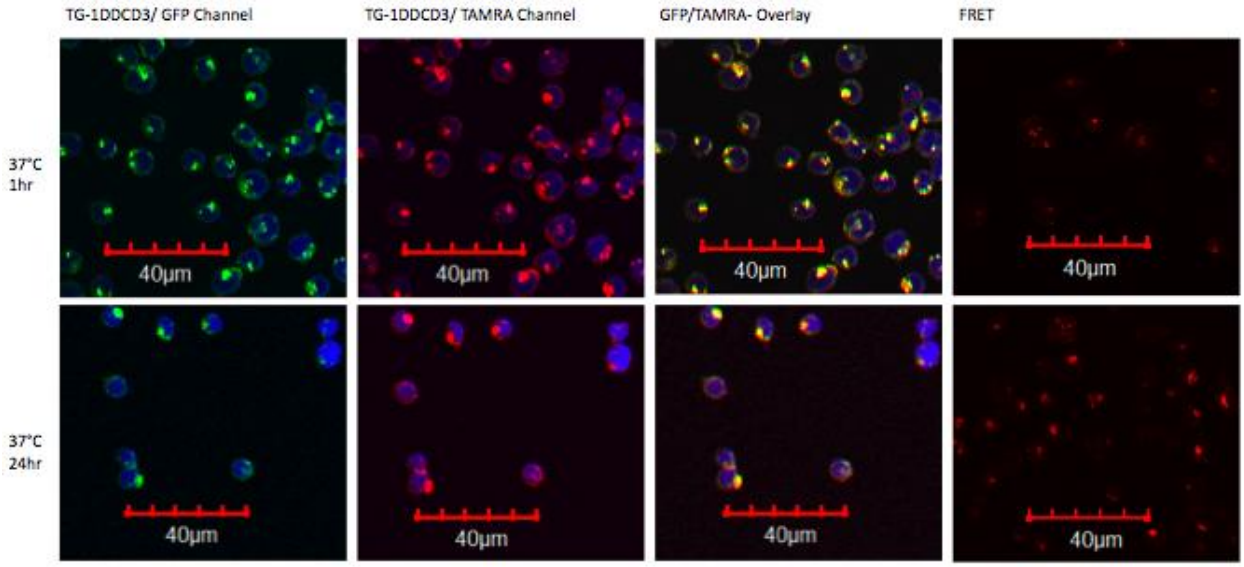


Figure 5.S14. Imaging of HPB-MLT cells treated with **TG-1DDCD3** at 37 °C for 1 h (top panel) and for 24 h (bottom panel). A & E: GFP channel; B & F: TAMRA channel; C &G: Overlay of the GFP and TAMRA channel indicates intact assembled protein (co-localized yellow punctuates); D & H: Observed FRET between GFP and TAMRA using above-mentioned filter settings. (FRET images were acquired right after GFP and TAMRA imaging and shows a different area of the same slide) Magnification: 40x.

Flow cytometry

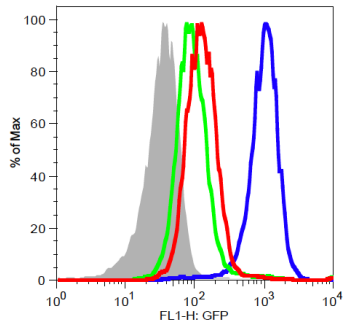


Figure 5.S15. Determination of CD3 receptor binding ability of assembled protein nanostructures in comparison to the anti-CD3 monoclonal antibody (Grey: untreated HPB-MLT cells; Red: HPB-MLT cells treated with 1 µM **TG-13DDCD3**; Green: HPB-MLT cells treated with 1 µM **TG-1DDCD3**; Blue- HPB-MLT cells treated with FITC labeled anti-CD3 monoclonal antibody).

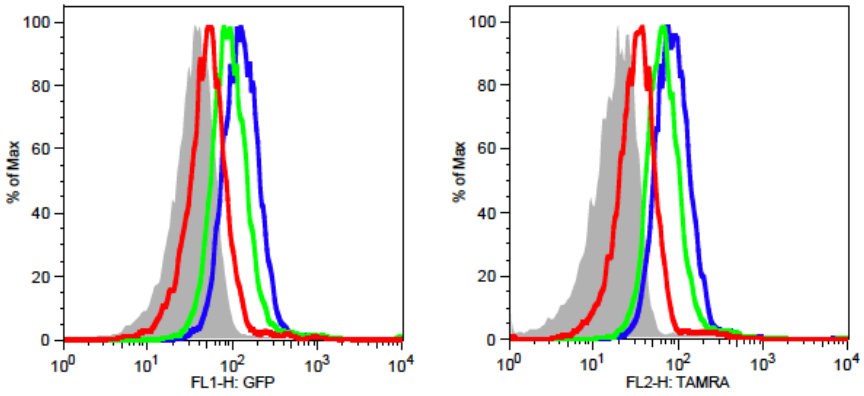
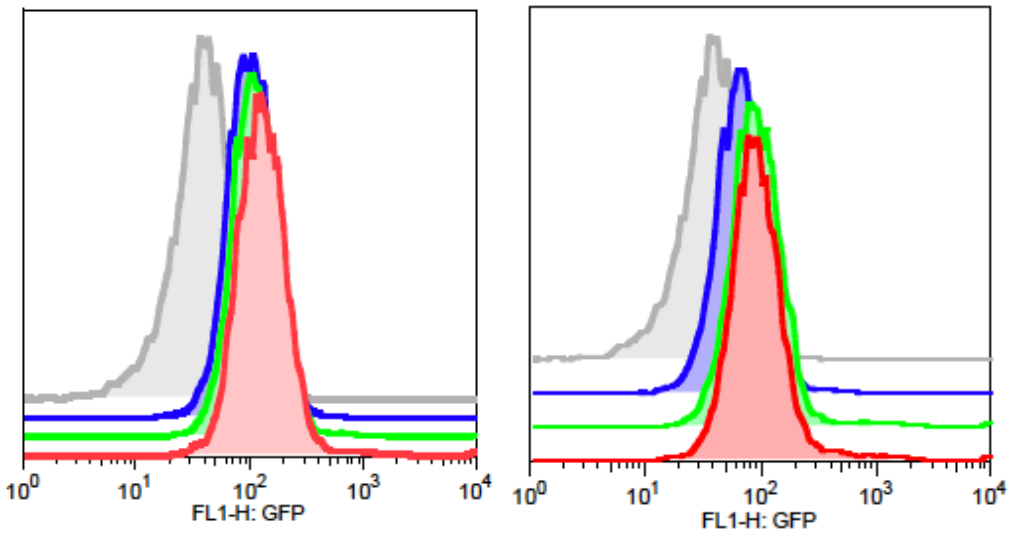
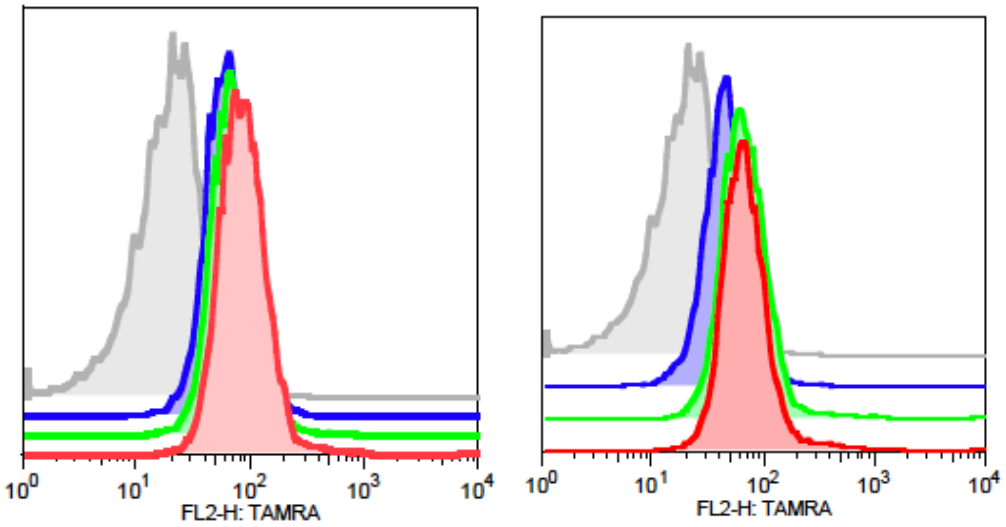


Figure 5.S16. Determination of CD3-specific binding of the assembled protein nanostructures to the HPB-MLT CD3 receptors (Left: GFP channel, Right: TAMRA channel). Grey: untreated HPB-MLT cells; Red: HPB-MLT cells treated with 1 µM **bisMTX-GFP-TAMRA** (**7**); Green: HPB-MLT cells treated with 1 µM **TG-1DDCD3**; Blue: HPB-MLT cells treated with 1 µM **TG-13DDCD3**.

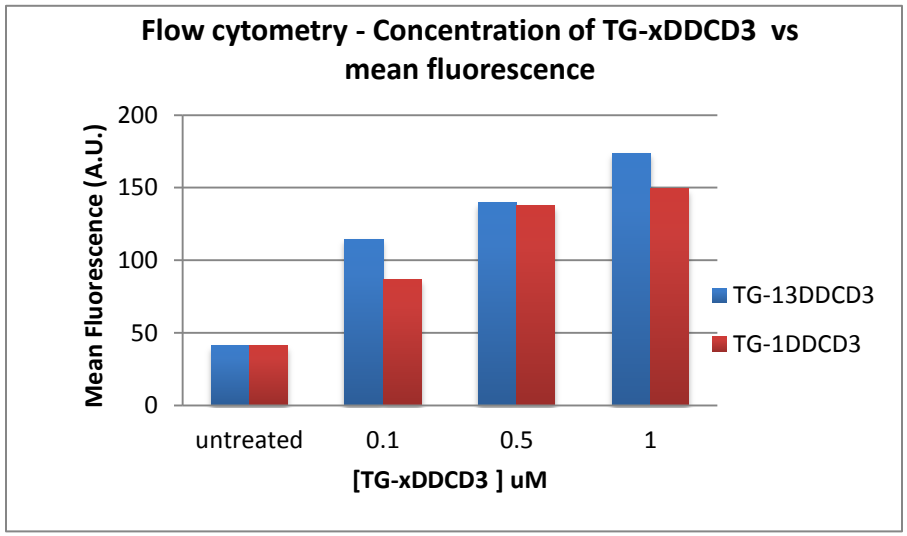


A **B**

Figure 5.S17. Determination of the dose dependent CD3 receptor binding of assembled protein nanostructures (GFP channel); A) HPB-MLT cells treated with increasing concentrations of **TG-13DDCD3** and B) with **TG-1DDCD3** (grey: untreated; blue: 0.1 µM; green: 0.5 µM; red: 1 µM).



A **B**



C

Figure 5.S18. Determination of the dose dependent CD3 receptor binding of assembled protein nanostructures (TAMRA channel); A) HPB-MLT cells treated with increasing concentrations of **TG-13DDCD3** and B) with **TG-1DDCD3** (grey: untreated; blue: 0.1 µM; green: 0.5 µM; red: 1 µM); C) Comparison of the **TG-13DDCD3** and **TG-1DDCD3** dose dependence using mean fluorescence values obtained from TAMRA channel.

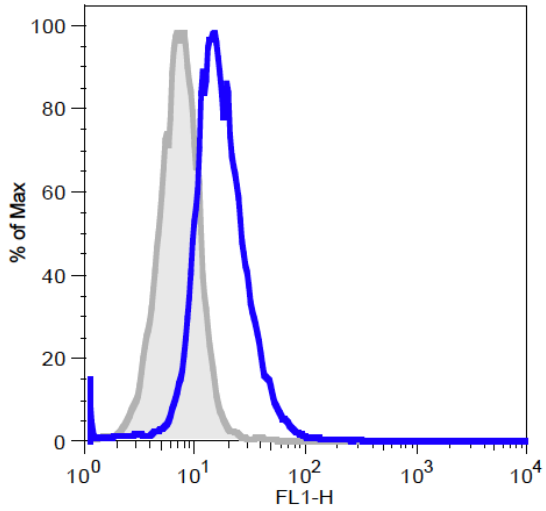
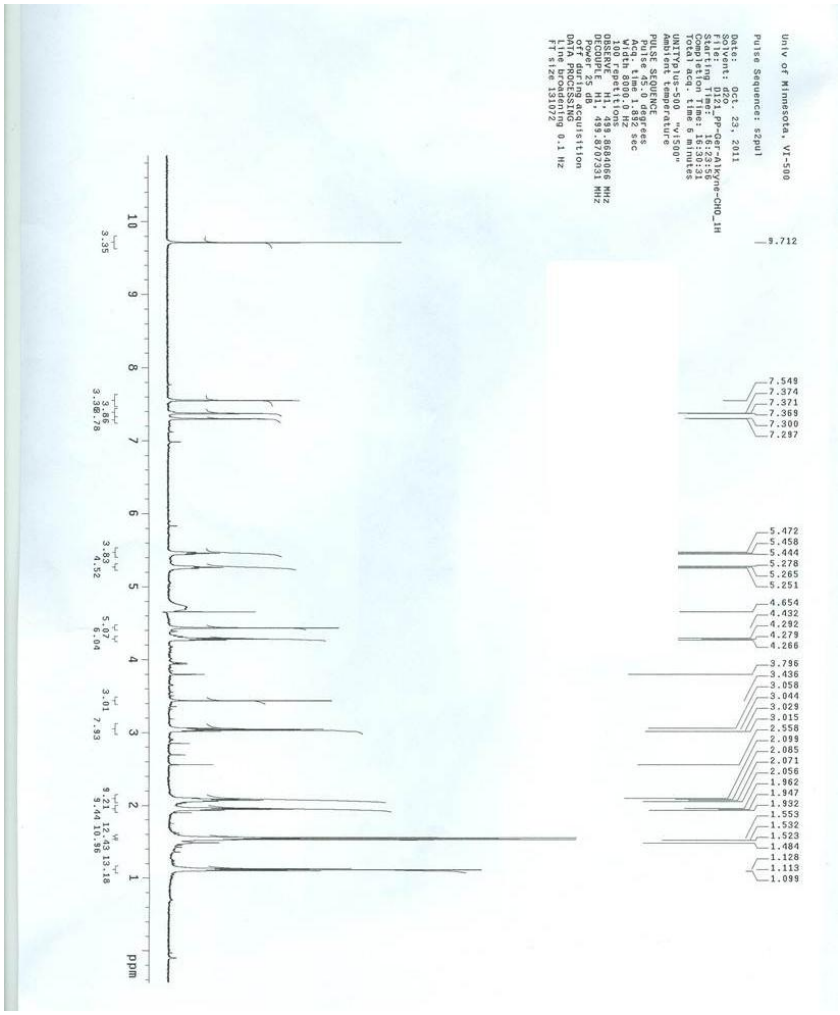
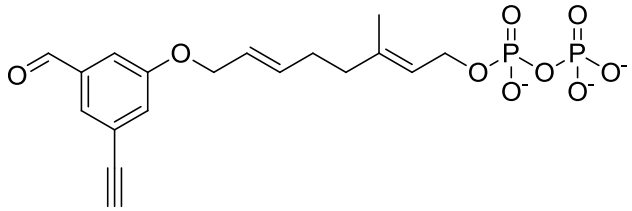


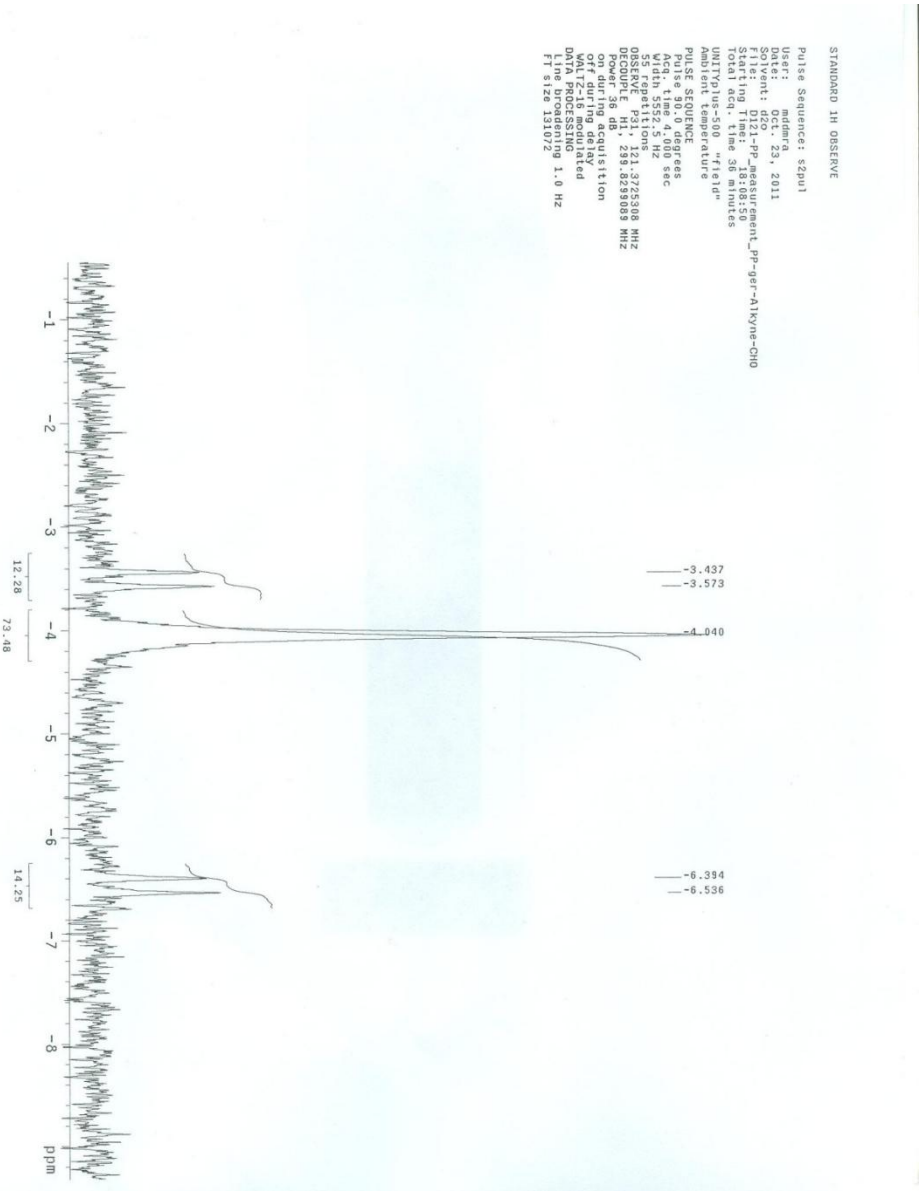
Figure 5.S19. Determination of the binding specificity of the assembled protein nanostructures: CD3 negative Daudi B cells were with treated with 1 µM **TG-13DDCD3** (grey: untreated cells; blue: **TG-13DDCD3** treated $CD3^-$ Daudi B lymphoma cells at 4 °C (2 h)).

NMRs.

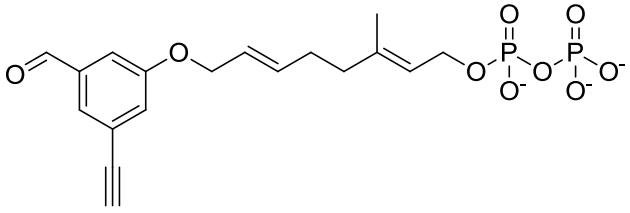


^{1}H-NMR of compound **1**





^{31}P-NMR of compound **1** containing H_3PO_4 as an internal standard**.**



References:

(1) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074.
(2) Rabuka, D. *Curr. Opin. Chem. Biol.* **2010**, *14*, 790–796.
(3) Carrico, I. S. *Chem. Soc. Rev.* **2008**, *37*, 1423.
(4) Wong, S. S.; Jameson, D. M.; Wong, S. S. *Chemistry of protein and nucleic acid cross-linking and conjugation*; Taylor & Francis/CRC Press: Boca Raton, 2012.
(5) Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M. *Adv. Drug. Deliv. Rev.* **2008**, *60*, 13–28.
(6) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Lee Wong, W.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W. *Nat. Biotechnol.* **2008**, *26*, 925–932.
(7) Gauchet, C.; Labadie, G. R.; Poulter, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 9274–9275.
(8) Kochendoerfer, G. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 555–560.
(9) Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. *Nat. Meth.* **2005**, *2*, 99–104.
(10) Köhn, M. *J. Pept. Sci.* **2009**, *15*, 393–397.
(11) Wong, L. S.; Khan, F.; Micklefield, J. *Chem. Rev.* **2009**, *109*, 4025–4053.
(12) Tiefenbrunn, T. K.; Dawson, P. E. *Biopolymers* **2010**, *94*, 95–106.
(13) Wu, B.-Y.; Hou, S.-H.; Huang, L.; Yin, F.; Zhao, Z.-X.; Anzai, J.-I.; Chen, Q. *Mater. Sci. Eng., C* **2008**, *28*, 1065–1069.
(14) Hermanson, G. T. *Bioconjugate techniques*; Academic Press: Amsterdam, 2008.
(15) Dixon, H. B. *Biochem. J.* **1964**, *92*, 661–666.
(16) Wu, P.; Brand, L. *Meth. Enzymol.* **1997**, *278*, 321–330.
(17) Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 11762–11770.
(18) Geoghegan, K. F.; Stroh, J. G. *Bioconjug. Chem.* **1992**, *3*, 138–146.
(19) Akgul, C.; Moulding, D. A.; White, M. R. .; Edwards, S. W. *FEBS Lett.* **2000**, *478*, 72–76.
(20) Dundr, M.; McNally, J. G.; Cohen, J.; Misteli, T. *J. Struc. Biol.* **2002**, *140*, 92–99.
(21) Kanda, T.; Sullivan, K. F.; Wahl, G. M. *Curr. Biol.* **1998**, *8*, 377–385.
(22) De Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E. *Bioconjugate Chem.* **2009**, *20*, 1281–1295.
(23) Wang, L.; Xie, J.; Schultz, P. G. *Annu Rev Biophys Biomol Struct* **2006**, *35*, 225–249.
(24) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055.

(25) Guo, J.; Melancon, C. E.; Lee, H. S.; Groff, D.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 9148–9151.
(26) Chatterjee, A.; Xiao, H.; Schultz, P. G. *Proc. Natl. Acad. Sci.* **2012**, *109*, 14841–14846.
(27) Huisgen, R.; Szeimies, G.; Mobius, L. *Chemische Berichte* **1967**, *100*, 2494.
(28) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
(29) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
(30) Köhn, M.; Breinbauer, R. *Angew. Chem. Int. Ed.* **2004**, *43*, 3106–3116.
(31) Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q. *Angew. Chem. Int. Ed.* **2008**, *47*, 2832–2835.
(32) Song, W.; Wang, Y.; Qu, J.; Lin, Q. *J. Am. Chem. Soc.* **2008**, *130*, 9654–9655.
(33) De Araújo, A. D.; Palomo, J. M.; Cramer, J.; Seitz, O.; Alexandrov, K.; Waldmann, H. *Chem. Eur. J.* **2006**, *12*, 6095–6109.
(34) De Araújo, A. D.; Palomo, J. M.; Cramer, J.; Köhn, M.; Schröder, H.; Wacker, R.; Niemeyer, C.; Alexandrov, K.; Waldmann, H. *Angew. Chem. Int. Ed.* **2006**, *45*, 296–301.
(35) Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y. *J. Am. Chem. Soc.* **2012**, *134*, 792–795.
(36) Blackman, M. L.; Royzen, M.; Fox, J. M. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
(37) Schoch, J.; Wiessler, M.; Jäschke, A. *J. Am. Chem. Soc.* **2010**, *132*, 8846–8847.
(38) Cordes, E. H.; Jencks, W. P. *J. Am. Chem. Soc.* **1962**, *84*, 826–831.
(39) Dirksen, A.; Yegneswaran, S.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2010**, 2023–2027.
(40) Rashidian, M.; Song, J. M.; Pricer, R. E.; Distefano, M. D. *J. Am. Chem. Soc.* **2012**, *134*, 8455–8467.
(41) Miller, L. W.; Cornish, V. W. *Curr. Opin. Chem. Biol.* **2005**, *9*, 56–61.
(42) Sunbul, M.; Yin, J. *Org. Biomol. Chem.* **2009**, *7*, 3361.
(43) Roeser, D.; Preusser-Kunze, A.; Schmidt, B.; Gasow, K.; Wittmann, J. G.; Dierks, T.; Von Figura, K.; Rudolph, M. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 81–86.
(44) Rush, J. S.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241.
(45) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2009**, *106*, 3000–3005.
(46) Rabuka, D.; Rush, J. S.; deHart, G. W.; Wu, P.; Bertozzi, C. R. *Nat. Protoc.* **2012**, *7*, 1052–1067.
(47) Yu, H.; Chokhawala, H.; Karpel, R.; Yu, H.; Wu, B.; Zhang, J.; Zhang, Y.; Jia, Q.; Chen, X. *J. Am. Chem. Soc.* **2005**, *127*, 17618–17619.
(48) Yu, H.; Huang, S.; Chokhawala, H.; Sun, M.; Zheng, H.; Chen, X. *Angew. Chem. Int. Ed.* **2006**, *45*, 3938–3944.
(49) Yin, J.; Liu, F.; Li, X.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126*, 7754–7755.

(50) Clarke, K. M.; Mercer, A. C.; La Clair, J. J.; Burkart, M. D. *J. Am. Chem. Soc.* **2005**, *127*, 11234–11235.
(51) Worthington, A. S.; Burkart, M. D. *Org. Biomol. Chem.* **2006**, *4*, 44.
(52) George, N.; Pick, H.; Vogel, H.; Johnsson, N.; Johnsson, K. *J. Am. Chem. Soc.* **2004**, *126*, 8896–8897.
(53) Cappellaro, C.; Baldermann, C.; Rachel, R.; Tanner, W. *EMBO J.* **1994**, *13*, 4737–4744.
(54) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15815–15820.
(55) Wong, L. S.; Thirlway, J.; Micklefield, J. *J. Am. Chem. Soc.* **2008**, *130*, 12456–12464.
(56) Vocadlo, D. J.; Hang, H. C.; Kim, E.; Hanover, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2003**, *100*, 9116–9121.
(57) Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.; Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2003**, *125*, 16162–16163.
(58) Clark, P. M.; Dweck, J. F.; Mason, D. E.; Hart, C. R.; Buck, S. B.; Peters, E. C.; Agnew, B. J.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2008**, *130*, 11576–11577.
(59) Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *Proc. Natl. Acad. Sci.* **2004**, *101*, 13132–13137.
(60) Khidekel, N.; Ficarro, S. B.; Clark, P. M.; Bryan, M. C.; Swaney, D. L.; Rexach, J. E.; Sun, Y. E.; Coon, J. J.; Peters, E. C.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2007**, *3*, 339–348.
(61) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science* **1999**, *285*, 760–763.
(62) Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009**, *10*, 787–798.
(63) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* **2007**, *3*, 707–708.
(64) Tanaka, T.; Yamamoto, T.; Tsukiji, S.; Nagamune, T. *ChemBioChem* **2008**, *9*, 802–807.
(65) Strijbis, K.; Spooner, E.; Ploegh, H. L. *Traffic* **2012**, *13*, 780–789.
(66) Möhlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. *ChemBioChem* **2011**, *12*, 1774–1780.
(67) Popp, M. W.; Dougan, S. K.; Chuang, T.-Y.; Spooner, E.; Ploegh, H. L. *Proc. Natl. Acad. Sci.* **2011**, *108*, 3169–3174.
(68) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J. Am. Chem. Soc.* **2008**, *130*, 16338–16343.
(69) Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J. Am. Chem. Soc.* **2009**, *131*, 9878–9879.
(70) Wu, Z.; Guo, X.; Guo, Z. *Chem. Commun.* **2011**, *47*, 9218.
(71) Sinisi, A.; Popp, M. W.-L.; Antos, J. M.; Pansegrau, W.; Savino, S.; Nissum, M.; Rappuoli, R.; Ploegh, H. L.; Buti, L. *Bioconjugate Chem.* **2012**, *23*, 1119–1126.

(72) Witte, M. D.; Cragnolini, J. J.; Dougan, S. K.; Yoder, N. C.; Popp, M. W.; Ploegh, H. L. *Proc. Natl. Acad. Sci.* **2012**, *109*, 11993–11998.
(73) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew. Chem. Int. Ed.* **2012**, *124*, 9511–9514.
(74) Jeger, S.; Zimmermann, K.; Blanc, A.; Grünberg, J.; Honer, M.; Hunziker, P.; Struthers, H.; Schibli, R. *Angew. Chem. Int. Ed.* **2010**, *49*, 9995–9997.
(75) Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A. *Bioconjugate Chem.* **2009**, *20*, 384–389.
(76) Abe, H.; Goto, M.; Kamiya, N. *Chem. Eur. J.* **2011**, *17*, 14004–14008.
(77) Tanaka, T.; Kamiya, N.; Nagamune, T. *FEBS Lett.* **2005**, *579*, 2092–2096.
(78) Tominaga, J.; Kamiya, N.; Doi, S.; Ichinose, H.; Maruyama, T.; Goto, M. *Biomacromolecules* **2005**, *6*, 2299–2304.
(79) Lin, C.-W.; Ting, A. Y. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.
(80) Duckworth, B. P.; Xu, J.; Taton, T. A.; Guo, A.; Distefano, M. D. *Bioconjugate Chem.* **2006**, *17*, 967–974.
(81) Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D. *ChemBioChem* **2007**, *8*, 98–105.
(82) Rashidian, M.; Dozier, J. K.; Lenevich, S.; Distefano, M. D. *Chem. Commun.* **2010**, *46*, 8998.
(83) Labadie, G. R.; Viswanathan, R.; Poulter, C. D. *J. Org. Chem.* **2007**, *72*, 9291–9297.
(84) Weinrich, D.; Lin, P.-C.; Jonkheijm, P.; Nguyen, U. T. T.; Schröder, H.; Niemeyer, C. M.; Alexandrov, K.; Goody, R.; Waldmann, H. *Angew. Chem. Int. Ed.* **2010**, *49*, 1252–1257.
(85) Nguyen, U. T. T.; Cramer, J.; Gomis, J.; Reents, R.; Gutierrez-Rodriguez, M.; Goody, R. S.; Alexandrov, K.; Waldmann, H. *ChemBioChem* **2007**, *8*, 408–423.
(86) Rose, M. W.; Xu, J.; Kale, T. A.; O'Doherty, G.; Barany, G.; Distefano, M. D. *Biopolymers* **2005**, *80*, 164–171.
(87) Rose, M. W.; Rose, N. D.; Boggs, J.; Lenevich, S.; Xu, J.; Barany, G.; Distefano, M. D. *J. Pept. Res.* **2005**, *65*, 529–537.
(88) Xu, J.; DeGraw, A. J.; Duckworth, B. P.; Lenevich, S.; Tann, C.-M.; Jenson, E. C.; Gruber, S. J.; Barany, G.; Distefano, M. D. *Chem. Biol. Drug. Des.* **2006**, *68*, 85–96.
(89) Duckworth, B. P.; Chen, Y.; Wollack, J. W.; Sham, Y.; Mueller, J. D.; Taton, T. A.; Distefano, M. D. *Angew. Chem.* **2007**, *119*, 8975–8978.
(90) Khatwani, S. L.; Kang, J. S.; Mullen, D. G.; Hast, M. A.; Beese, L. S.; Distefano, M. D.; Taton, T. A. *Bioorg. Med. Chem.* **2012**, *20*, 4532–4539.
(91) Speers, A. E.; Cravatt, B. F. *Chem. Biol.* **2004**, *11*, 535–546.
(92) Hosokawa, A.; Wollack, J. W.; Zhang, Z.; Chen, L.; Barany, G.; Distefano, M. D. *Int. J. Pept. Res. Ther.* **2007**, *13*, 345–354.
(93) Wollack, J. W.; Silverman, J. M.; Petzold, C. J.; Mougous, J. D.; Distefano, M. D. *ChemBioChem* **2009**, *10*, 2934–2943.

(94) Algar, W. R.; Prasuhn, D. E.; Stewart, M. H.; Jennings, T. L.; Blanco-Canosa, J. B.; Dawson, P. E.; Medintz, I. L. *Bioconjugate Chem.* **2011**, *22*, 825–858.
(95) Howarth, M. *Proc. Natl. Acad. Sci.* **2005**, *102*, 7583–7588.
(96) Howarth, M.; Liu, W.; Puthenveetil, S.; Zheng, Y.; Marshall, L. F.; Schmidt, M. M.; Wittrup, K. D.; Bawendi, M. G.; Ting, A. Y. *Nat. Meth.* **2008**, *5*, 397–399.
(97) Slavoff, S. A.; Chen, I.; Choi, Y.-A.; Ting, A. Y. *J. Am. Chem. Soc.* **2008**, *130*, 1160–1162.
(98) Fernández-Suárez, M.; Chen, T. S.; Ting, A. Y. *J. Am. Chem. Soc.* **2008**, *130*, 9251–9253.
(99) Sueda, S.; Yoneda, S.; Hayashi, H. *ChemBioChem* **2011**, *12*, 1367–1375.
(100) Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.
(101) Cohen, J. D.; Thompson, S.; Ting, A. Y. *Biochemistry* **2011**, *50*, 8221–8225.
(102) Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernandez-Suarez, M.; Puthenveetil, S.; Ting, A. Y. *Proc. Natl. Acad. Sci.* **2010**, *107*, 10914–10919.
(103) Slavoff, S. A.; Liu, D. S.; Cohen, J. D.; Ting, A. Y. *J. Am. Chem. Soc.* **2011**, *133*, 19769–19776.
(104) Cohen, J. D.; Zou, P.; Ting, A. Y. *ChemBioChem* **2012**, *13*, 888–894.
(105) Heal, W. P.; Wickramasinghe, S. R.; Bowyer, P. W.; Holder, A. A.; Smith, D. F.; Leatherbarrow, R. J.; Tate, E. W. *Chem. Comm.* **2008**, 480.
(106) Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6*, 2308.
(107) Mu, Y.; Gibbs, R. A.; Eubanks, L. M.; Poulter, C. D. *J. Org. Chem.* **1996**, *61*, 8010–8015.
(108) Subramanian, T.; Liu, S.; Troutman, J. M.; Andres, D. A.; Spielmann, H. P. *ChemBioChem* **2008**, *9*, 2872–2882.
(109) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. *Science* **2001**, *293*, 2101–2105.
(110) Luk, Y.-Y.; Tingey, M. L.; Dickson, K. A.; Raines, R. T.; Abbott, N. L. *J. Am. Chem. Soc.* **2004**, *126*, 9024–9032.
(111) Weinrich, D.; Lin, P.-C.; Jonkheijm, P.; Nguyen, U. T. T.; Schröder, H.; Niemeyer, C. M.; Alexandrov, K.; Goody, R.; Waldmann, H. *Angew. Chem. Int. Ed.* **2010**, *49*, 1252–1257.
(112) Zhong, M.; Fang, J.; Wei, Y. *Bioconjugate Chem.* **2010**, *21*, 1177–1182.
(113) Lin, P.-C.; Ueng, S.-H.; Tseng, M.-C.; Ko, J.-L.; Huang, K.-T.; Yu, S.-C.; Adak, A. K.; Chen, Y.-J.; Lin, C.-C. *Angew. Chem. Int. Ed.* **2006**, *45*, 4286–4290.
(114) Goldstein, D. C.; Thordarson, P.; Peterson, J. R. *Aust. J. Chem.* **2009**, *62*, 1320.
(115) Lin, P.-C.; Weinrich, D.; Waldmann, H. *Macromol. Chem. Phys.* **2010**, *211*, 136–144.
(116) Yamaguchi, H.; Miyazaki, M.; Honda, T.; Briones-Nagata, M. P.; Arima, K.; Maeda, H. *Electrophoresis* **2009**, *30*, 3257–3264.

(117) Lee, Y.; Lee, E. K.; Cho, Y. W.; Matsui, T.; Kang, I.-C.; Kim, T.-S.; Han, M. H. *Proteomics* **2003**, *3*, 2289–2304.
(118) Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K. *Proc. Natl. Acad. Sci.* **2004**, *101*, 9955–9959.
(119) Chen, I.; Ting, A. Y. *Curr. Opin. Biotechnol.* **2005**, *16*, 35–40.
(120) Torchilin, V. P.; Lukyanov, A. N. *Drug Discov. Today* **2003**, *8*, 259–266.
(121) Discher, D. E. *Science* **2002**, *297*, 967–973.
(122) Rizzi, S. C.; Hubbell, J. A. *Biomacromolecules* **2005**, *6*, 1226–1238.
(123) Kurpiers, T.; Mootz, H. D. *Angew. Chem. Int. Ed.* **2009**, *48*, 1729–1731.
(124) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347.
(125) Blackman, M. L.; Royzen, M.; Fox, J. M. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
(126) Codelli, J. A.; Baskin, J. M.; Agard, N. J.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2008**, *130*, 11486–11493.
(127) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2007**, *104*, 16793–16797.
(128) Lallana, E.; Fernandez-Megia, E.; Riguera, R. *J. Am. Chem. Soc.* **2009**, *131*, 5748–5750.
(129) Dirksen, A.; Dirksen, S.; Hackeng, T. M.; Dawson, P. E. *J. Am. Chem. Soc.* **2006**, *128*, 15602–15603.
(130) Zeng, Y.; Ramya, T. N. C.; Dirksen, A.; Dawson, P. E.; Paulson, J. C. *Nat Meth* **2009**, *6*, 207–209.
(131) Rayo, J.; Amara, N.; Krief, P.; Meijler, M. M. *J. Am. Chem. Soc.* **2011**, *133*, 7469–7475.
(132) Dixon, H. B. F. *J. Protein Chem.* **1984**, *3*, 99–108.
(133) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 5307–5311.
(134) Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66.
(135) Turek, T. C.; Gaon, I.; Distefano, M. D.; Strickland, C. L. *J. Org. Chem.* **2001**, *66*, 3253–3264.
(136) Yakhnin, A. V.; Vinokurov, L. M.; Surin, A. K.; Alakhov, Y. B. *Protein Expres. Pur.* **1998**, *14*, 382–386.
(137) Inagaki, N.; Seino, Y.; Takeda, J.; Yano, H.; Yamada, Y.; Bell, G. I.; Eddy, R. L.; Fukushima, Y.; Byers, M. G.; Shows, T. B.; Imura, H. *Molecular Endocrinology* **1989**, *3*, 1014–1021.
(138) Gault, V.; Kerr, B.; Irwin, N.; Flatt, P. *Biochem Pharmacol.* **2008**, *75*, 2325–2333.
(139) Gaon, I.; Turek, T.; Distefano, M. *Tetrahedron Lett.* **1996**, *37*, 8833–8836.
(140) Lenevich, S.; Distefano, M. D. *Analytical Biochem.* **2011**, *408*, 316–320.
(141) Kalia, J.; Raines, R. T. *Angew. Chem. Int. Ed.* **2008**, *47*, 7523–7526.

(142) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2006**, *45*, 7581–7584.
(143) Jevsevar, S.; Kunstelj, M.; Porekar, V. G. *Biotechnol J.* **2010**, *5*, 113–128.
(144) Veronese, F. M.; Mero, A.; Pasut, G. In *PEGylated Protein Drugs: Basic Science and Clinical Applications*; Veronese, F. M., Ed.; Birkhäuser Basel: Basel, 2009; pp. 11–31.
(145) Veronese, F. M. *Biomaterials* **2001**, *22*, 405–417.
(146) Roberts, M. J.; Bentley, M. D.; Harris, J. M. *Advanced Drug Delivery Reviews* **2002**, *54*, 459–476.
(147) Flatt, P. R. *Diabet Med* **2008**, *25*, 759–764.
(148) Elahi, D.; McAloon-Dyke, M.; Fukagawa, N. K.; Meneilly, G. S.; Sclater, A. L.; Minaker, K. L.; Habener, J. F.; Andersen, D. K. *Regul. Pept.* **1994**, *51*, 63–74.
(149) Friday, B. B.; Adjei, A. A. *Biochim Biophys Acta.* **2005**, *1756*, 127–144.
(150) Maynard, H. D.; Broyer, R. M.; Kolodziej, C. M. In *Click Chemistry for Biotechnology and Materials Science*; Lahann, J., Ed.; John Wiley & Sons, Ltd: Chichester, UK, 2009; pp. 53–68.
(151) Singh, Y.; Renaudet, O.; Defrancq, E.; Dumy, P. *Org. Lett.* **2005**, *7*, 1359–1362.
(152) Sukhorukov, A. Y.; Ioffe, S. L. *Chem. Rev.* **2011**, *111*, 5004–5041.
(153) Dirksen, A.; Dawson, P. E. *Bioconjugate Chem.* **2008**, *19*, 2543–2548.
(154) Park, S.; Yousaf, M. N. *Langmuir* **2008**, *24*, 6201–6207.
(155) Von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Nat Chem* **2010**, *2*, 96–101.
(156) Lempens, E. H. M.; Helms, B. A.; Merkx, M.; Meijer, E. W. *ChemBioChem* **2009**, *10*, 658–662.
(157) Sander, E. G.; Jencks, W. P. *J. Am. Chem. Soc.* **1968**, *90*, 6154–6162.
(158) Baskin, J. M.; Dehnert, K. W.; Laughlin, S. T.; Amacher, S. L.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2010**, *107*, 10360–10365.
(159) Pompliano, D. L.; Gomez, R. P.; Anthony, N. J. *J. Am. Chem. Soc.* **1992**, *114*, 7945–7946.
(160) Cordes, E. H.; Jencks, W. P. *J. Am. Chem. Soc.* **1962**, *84*, 832–837.
(161) Wen-jun, W.; Chen-ming, C.; Chen, J.; Xin, W.; George-peng, W. *Chem. Res. Chinese Universities* **2011**, *27*, 886–890.
(162) Thom, J.; Anderson, D.; McGregor, J.; Cotton, G. *Bioconjug. Chem.* **2011**, *22*, 1017–1020.
(163) Ip, N. Y.; Yancopoulos, G. D. *Annu. Rev. Neurosci.* **1996**, *19*, 491–515.
(164) Rhee, K. D.; Yang, X.-J. *Adv. Exp. Med. Biol.* **2010**, *664*, 647–654.
(165) Wen, R.; Tao, W.; Li, Y.; Sieving, P. A. *Prog Retin Eye Res* **2012**, *31*, 136–151.
(166) Fleissner, M. R.; Brustad, E. M.; Kalai, T.; Altenbach, C.; Cascio, D.; Peters, F. B.; Hideg, K.; Peuker, S.; Schultz, P. G.; Hubbell, W. L. *Proceedings of the National Academy of Sciences* **2009**, *106*, 21637–21642.
(167) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. *J. Am. Chem. Soc.* **2012**, 120606110334009.
(168) Yi, L.; Sun, H.; Wu, Y.-W.; Triola, G.; Waldmann, H.; Goody, R. S. *Angew. Chem. Int. Ed.* **2010**, *49*, 9417–9421.

(169) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.
(170) Weissleder, R. *Science* **2006**, *312*, 1168–1171.
(171) Schlemmer, H.-P. W.; Pichler, B. J.; Schmand, M.; Burbar, Z.; Michel, C.; Ladebeck, R.; Jattke, K.; Townsend, D.; Nahmias, C.; Jacob, P. K.; Heiss, W.-D.; Claussen, C. D. *Radiology* **2008**, *248*, 1028–1035.
(172) Nahrendorf, M.; Zhang, H.; Hembrador, S.; Panizzi, P.; Sosnovik, D. E.; Aikawa, E.; Libby, P.; Swirski, F. K.; Weissleder, R. *Circulation* **2008**, *117*, 379–387.
(173) Gaon, I.; Turek, T. C.; Weller, V. A.; Edelstein, R. L.; Singh, S. K.; Distefano, M. D. *J. Org. Chem.* **1996**, *61*, 7738–7745.
(174) Shirayev, A.; Thoo lin, P. K.; Moiseev, I. K. *Synthesis* **1997**, *1997*, 38–40.
(175) Satyanarayana, M.; Vitali, F.; Frost, J. R.; Fasan, R. *Chem. Commun. (Camb.)* **2012**, *48*, 1461–1463.
(176) Li, Q.; Hapka, D.; Chen, H.; Vallera, D. A.; Wagner, C. R. *Angew. Chem. Int. Ed.* **2008**, *47*, 10179–10182.
(177) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361–374.
(178) Boder, E. T.; Midelfort, K. S.; Wittrup, K. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10701–10705.
(179) Jaeger, K.-E.; Eggert, T. *Curr. Opin. Biotechnol.* **2004**, *15*, 305–313.
(180) Kolkman, J. A.; Stemmer, W. P. *Nat. Biotechnol.* **2001**, *19*, 423–428.
(181) Kuchner, O.; Arnold, F. H. *Trends Biotechnol.* **1997**, *15*, 523–530.
(182) Shi, J.; Xiao, Z.; Kamaly, N.; Farokhzad, O. C. *Acc. Chem. Res.* **2011**, *44*, 1123–1134.
(183) Jutz, G.; Böker, A. *Polymer* **2011**, *52*, 211–232.
(184) Rothemund, P. W. K. *Nature* **2006**, *440*, 297–302.
(185) Pistol, C.; Dwyer, C. *Nanotechnology* **2007**, *18*, 125305.
(186) Guo, P. *Nat. Nanotechnol.* **2010**, *5*, 833–842.
(187) Ke, Y.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H. *Science* **2008**, *319*, 180–183.
(188) Matsuura, K.; Watanabe, K.; Matsuzaki, T.; Sakurai, K.; Kimizuka, N. *Angew. Chem. Int. Ed.* **2010**, *49*, 9662–9665.
(189) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. *Acc. Chem. Res.* **2010**, *43*, 661–672.
(190) Hu, M.; Qian, L.; Briñas, R. P.; Lymar, E. S.; Hainfeld, J. F. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5111–5114.
(191) Lai, J. R.; Fischbach, M. A.; Liu, D. R.; Walsh, C. T. *J. Am. Chem. Soc.* **2006**, *128*, 11002–11003.
(192) Chen, G.; Jiang, M. *Chem. Soc. Rev.* **2011**, *40*, 2254.
(193) Uhlenheuer, D. A.; Petkau, K.; Brunsveld, L. *Chem. Soc. Rev.* **2010**, *39*, 2817.
(194) Fegan, A.; White, B.; Carlson, J. C. T.; Wagner, C. R. *Chem. Rev.* **2010**, *110*, 3315–3336.
(195) Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44*, 651–653.
(196) Prescher, J. A.; Bertozzi, C. R. *Nat Chem Biol* **2005**, *1*, 13–21.
(197) Hudak, J. E.; Barfield, R. M.; De Hart, G. W.; Grob, P.; Nogales, E.; Bertozzi, C. R.; Rabuka, D. *Angew. Chem., Int. Ed.* **2012**, *51*, 4161–4165.

(198) White, B. R.; Li, Q.; Wagner, C. R. *Methods Mol. Biol.* **2011**, *743*, 17–26.
(199) Carlson, J. C. T.; Jena, S. S.; Flenniken, M.; Chou, T.; Siegel, R. A.; Wagner, C. R. *J. Am. Chem. Soc.* **2006**, *128*, 7630–7638.
(200) Li, Q.; So, C. R.; Fegan, A.; Cody, V.; Sarikaya, M.; Vallera, D. A.; Wagner, C. R. *J. Am. Chem. Soc.* **2010**, *48*, 17247–17257.
(201) Gangar, A.; Fegan, A.; Kumarapperuma, S. C.; Wagner, C. R. *J. Am. Chem. Soc.* **2012**, *134*, 2895–2897.
(202) Placzek, A. T.; Gibbs, R. A. *Org. Lett.* **2011**, 3576–3579.
(203) Sonogashira, K. *J. Organomet. Chem.* **2002**, *653*, 46–49.
(204) Rashidian, M.; Mahmoodi, M. M.; Distefano, M. D. *Bioconjug. Chem.* **2012**, *Submitted*.
(205) Zbinden, K. G.; Obst-Sander, U.; Hilpert, K.; Kühne, H.; Banner, D. W.; Böhm, H.-J.; Stahl, M.; Ackermann, J.; Alig, L.; Weber, L.; Wessel, H. P.; Riederer, M. A.; Tschopp, T. B.; Lavé, T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5344–5352.
(206) Gaon, I.; Turek, T.; Distefano, M. *Tetrahedron Lett.* **1996**, *37*, 8833–8836.
(207) Fegan, A.; Kumarapperuma, S. C.; Wagner, C. R. *Molecular Pharmaceutics* **2012**, Submitted.